(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 4 July 2002 (04.07.2002)

PCT

(10) International Publication Number WO 02/052015 A2

- (51) International Patent Classification⁷: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/257.857

22 December 2000 (22.12.2000) US

- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S. 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

CLONE APGK-12			AMINO ACID SUBSTITUTIONS							
	5' HCV		EMCV		HC	V NS2→5B				3'HCV
	IRES	NeoR:	IRES	NS2	NS3	4A	NS4B	NS5A	NS5B	UTR
Ti chulug	G (nt1) SEQ ID NO 1									
	A (nt1) SEQ ID NO 24			-	•	-	-	-	•	
86 ch/µg	R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1891)R	.	T(1993)A G(2042)C L(2155)P P(2166)L		
2000000ctu/µg	G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1891)R	-	T(1993)A G(2042)C L(2155)P P(2166)L		

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.



0.02/052015



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

1

SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

FIELD OF THE INVENTION

The present invention relates generally to a HCV RNA molecule that self-replicates in appropriate cell lines, particularly to a self-replicating HCV RNA construct having an enhanced efficiency of establishing cell culture replication.

BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion and community-acquired non-A non-B hepatitis worldwide. It is estimated that over 200 million people worldwide are infected by the virus. A high percentage of carriers become chronically infected and many progress to chronic liver disease, so called chronic hepatitis C. This group is in turn at high risk for serious liver disease such as liver cirrhosis, hepatocellular carcinoma and terminal liver disease leading to death. The mechanism by which HCV establishes viral persistence and causes a high rate of chronic liver disease has not been thoroughly elucidated. It is not known how HCV interacts with and evades the host immune system. In addition, the roles of cellular and humoral immune responses in protection against HCV infection and disease have yet to be established.

20

25

35

15

Various clinical studies have been conducted with the goal of identifying pharmaceutical compounds capable of effectively treating HCV infection in patients afflicted with chronic hepatitis C. These studies have involved the use of interferonalpha, alone and in combination with other antiviral agents such as ribavirin. Such studies have shown that a substantial number of the participants do not respond to these therapies, and of those that do respond favorably, a large proportion were found to relapse after termination of treatment. To date there are no broadly effective antiviral compounds for treatment of HCV infection.

30 HCV is an enveloped positive strand RNA virus in the *Flaviviridae* family. The single strand HCV RNA genome is of positive polarity and comprises one open reading frame (ORF) of approximately 9600 nucleotides in length, which encodes a linear polyprotein of approx. 3010 amino acids. In infected cells, this polyprotein is cleaved

at multiple sites by cellular and viral proteases to produce structural and nonstructural (NS) proteins. The structural proteins (C, E1, E2 and E2-p7) comprise

2

polypeptides that constitute the virus particle (Hijikata et al., 1991; Grakoui et al., 1993(a)). The non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B) encode for enzymes or accessory factors that catalyze and regulate the replication of the HCV RNA genome. Processing of the structural proteins is catalyzed by host cell proteases (Hijikata et al., 1991). The generation of the mature non-structural proteins is catalyzed by two virally encoded proteases. The first is the NS2/3 zincdependent metalloprotease which auto-catalyses the release of the NS3 protein from the polyprotein. The released NS3 contains a N-terminal serine protease domain (Grakoui et al., 1993(b); Hijikata et al., 1993) and catalyzes the remaining cleavages from the polyprotein. The released NS4A protein has at least two roles, First. forming a stable complex with NS3 protein and assisting in the membrane localization of the NS3/NS4A complex (Kim et al., Arch Virol, 1999, 144; 329-343) and second, acting as a cofactor for NS3 protease activity. This membraneassociated complex, in turn catalyzes the cleavage of the remaining sites on the polyprotein, thus effecting the release of NS4B, NS5A and NS5B (Bartenschlager et al., 1993; Grakoui et al., 1993(a); Hijikata et al., 1993; Love et al., 1996; reviewed in Kwong et al., 1998). The C-terminal segment of the NS3 protein also harbors nucleoside triphosphatase and RNA helicase activity (Kim et al., 1995). The function of the protein NS4B is unknown. NS5A, a highly phosphorylated protein, seems to be responsible for the Interferon resistance of various HCV genotypes (Gale Jr. et al. 1997 Virology 230, 217; Reed et al., 1997. NS5B is an RNA-dependent RNA polymerase (RdRp) that is involved in the replication of HCV.

10

15

20

35

The open reading frame of the HCV RNA genome is flanked on its 5' end by a nontranslated region (NTR) of approx. 340 nucleotides that functions as the internal ribosome entry site (IRES), and on its 3' end by a NTR of approximately 230 nucleotides. Both the 5' and 3' NTRs are important for RNA genome replication. The genomic sequence variance is not evenly distributed over the genome and the 5'NTR and parts of the 3'NTR are the most highly conserved portions. The authentic, highly conserved 3'NTR is the object of US patent 5,874,565 granted to Rice et al.

The cloned and characterized partial and complete sequences of the HCV genome have also been analyzed with regard to appropriate targets for a prospective antiviral therapy. Four viral enzyme activities provide possible targets such as (1) the NS2/3 protease; (2) the NS3/4A protease complex, (3) the NS3 Helicase and (4) the NS5B

3

RNA-dependent RNA polymerase. The NS3/4A protease complex and the NS3 helicase have already been crystallized and their three-dimensional structure determined (Kim et al., 1996; Yem et al., 1998; Love et al., 1996; Kim et al., 1998; Yao et al., 1997; Cho et al., 1998). The NS5B RNA dependent RNA polymerase has also been crystallized to reveal a structure reminiscent of other nucleic acid polymerases (Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039; Ago et al. 1999, Structure 7: 1417-1426; Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943).

5

20

25

30

35

Even though important targets for the development of a therapy for chronic HCV infection have been defined with these enzymes and even though a worldwide intensive search for suitable inhibitors is ongoing with the aid of rational drug design and HTS, the development of therapy has one major deficiency, namely the lack of cell culture systems or simple animal models, which allow direct and reliable
 propagation of HCV viruses. The lack of an efficient cell culture system is still the main reason to date that an understanding of HCV replication remains elusive.

Although flavi- and pestivirus self-replicating RNAs have been described and used for the replication in different cell lines with a relatively high yield, similar experiments with HCV have not been successful to date (Khromykh *et al.*, 1997; Behrens *et al.*, 1998; Moser *et al.*, 1998). It is known from different publications that cell lines or primary cell cultures can be infected with high-titer patient serum containing HCV (Lanford *et al.* 1994; Shimizu *et al.* 1993; Mizutani *et al.* 1996; Ikda *et al.* 1998; Fourner *et al.* 1998; Ito *et al.* 1996). However, these virus-infected cell lines or cell cultures do not allow the direct detection of HCV-RNA or HCV antigens.

It is also known from the publications of Yoo et al. 1995; and of Dash et al., 1997; that hepatoma cell lines can be transfected with synthetic HCV-RNA obtained through *in vitro* transcription of the cloned HCV genome. In both publications the authors started from the basic idea that the viral HCV genome is a plus-strand RNA functioning directly as mRNA after being transfected into the cell, permitting the synthesis of viral proteins in the course of the translation process, and so new HCV particles could form HCV viruses and their RNA detected through RT-PCR. However the published results of the RT-PCR experiments indicate that the HCV replication in the described HCV transfected hepatoma cells is not particularly

5

10

4

efficient and not sufficient to measure the quality of replication, let alone measure the modulations in replication after exposure to potential antiviral drugs. Furthermore it is now known that the highly conserved 3' NTR is essential for the virus replication (Yanagi *et al.*, 1999). This knowledge strictly contradicts the statements of Yoo *et al.* (*supra*) and Dash *et al.* (*supra*), who used for their experiments only HCV genomes with shorter 3' NTRs and not the authentic 3' end of the HCV genome.

In WO 98/39031, Rice *et al.* disclosed authentic HCV genome RNA sequences, in particular containing: a) the highly conserved 5'-terminal sequence "GCCAGCC"; b) the HCV polyprotein coding region; and c) 3'-NTR authentic sequences.

In WO 99/04008, Purcell *et al.* disclosed an HCV infectious clone that also contained only the highly conserved 5'-terminal sequence "GCCAGC".

Recently Lohman *et al.* **1999** (Science 285: 110-113) and Bartenschlager *et al.* (in CA 2,303,526, laid-open on October 3, 2000) disclosed a HCV cell culture system where the viral RNA (I377/NS2-3') self-replicates in the transfected cells with such efficiency that the quality of replication can be measured with accuracy and reproducibility. The Lohman and Bartenschlager disclosures were the first demonstration of HCV RNA replication in cell culture that was substantiated through direct measurement by Northern blots. This replicon system and sequences disclosed therein highlight once again the conserved 5' sequence "GCCAGC". A similar observation highlighting the conservation of the 5'NTR was made by Blight *et al.* **2000** (Science 290: 1972-1974) and WO 01/89364 published on Nov. 29, 2001.

25

30

35

In addition to the conservation of the 5' and 3' untranslated regions in cell culture replicating RNAs, three other publications by Lohman *et al.* **2001**, Krieger *et al.* **2001** and Guo *et al.* **2001** have recently disclosed distinct adaptive mutants within the HCV non-structural protein coding region. Specific nucleotide changes that alter the amino acids of the HCV non-structural proteins are shown to enhance the efficiency of establishing stable replicating HCV subgenomic replicons in culture cells.

Applicant has now found that, contrary to all previous reports, the highly conserved 5'-NTR can be mutated by adaptation to give rise to a HCV RNA sequence that, in conjunction with mutations in the HCV non-structural region, provides for a greater

5

efficiency of transduction and/or replication.

Applicant has also identified novel adaptive mutations within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture.

One advantage of the present invention is to provide an alternative to these existing systems comprising a HCV RNA molecule that self-replicates. Moreover, the present invention demonstrates that the initiating nucleotide of the plus-strand genome can be either an A as an alternative to the G already disclosed.

A further advantage of the present invention is to provide a unique HCV RNA molecule that transduces and/or replicates with higher efficiency. The Applicant demonstrates the utility of this specific RNA molecule in a cell line and its use in evaluating a specific inhibitor of HCV replication.

SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides a 5'-non translated region of
the hepatitis C virus wherein its highly conserved guanine at position 1 is substituted
for adenine.

Particularly, the present invention provides a hepatitis C virus polynucleotide comprising adenine at position 1 as numbered according to the I377/NS2-3' construct (Lohmann et al. 1999, Accession # AJ242651).

Particularly, the invention provides a HCV self-replicating polynucleotide comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO. 8).

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

5

10

15

25

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

More particularly, the invention provides a HCV self-replicating polynucleotide encoding a polyprotein comprising a G2042C or a G2042R mutation.

Most particularly, the invention provides for HCV self-replicating polynucleotide comprising a nucleotide substitution G->A at position 1, and said polynucleotide encodes a polyprotein further comprising a G2042C or a G2042R mutation.

Particularly, the polynucleotide of the present invention can be in the form of RNA or DNA that can be transcribed to RNA.

In a third embodiment, the invention also provides for an expression vector comprising a DNA form of the above polynucleotide, operably linked with a promoter.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or the vector as described above.

20

10

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- incubating the host cell as described above in the absence or presence of a potential hepatitis C virus inhibitor;
- 25 isolating the total cellular RNA from the cells;
 - analyzing the RNA so as to measure the amount of HCV RNA replicated;
 - comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.
- In a sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) treating the above described host cell with the compound:
 - b) evaluating the treated host cell for reduced replication, wherein reduced replication indicates the ability of the compound to inhibit replication.

7

DETAILED DESCRIPTION OF THE DRAWINGS

10

Figure 1 is a schematic view of the bi-cistronic replicon RNA. The sequence deviations between the I377/NS2-3' replicon from Lohman *et al.*, 1999 and the APGK12 replicon are indicated below the replicon. In place of a G nucleotide at the +1 position in the I377/NS2-3' replicon, the APGK12 contains an additional G resulting in GG at the 5' terminus (the first G being counted as position –1). In the linker region between the neo gene and the EMCV IRES sequence two areas deviate from I377/NS2-3': 14 nucleotides (CGCGCCCAGATGTT) which are not present in I377/NS2/3 are inserted at position 1184 in APGK12; 11 nucleotides (1231-1241) present in I377/NS2-3' are deleted to generate APGK-12. In the NS5B coding region, a T at position 8032 was mutated to C to eliminate a Ncol restriction site.

- 15 Figure 2 shows Northern blots of RNA-transfected Huh-7 cell lines. 12 µg of total . cellular RNA or control RNA was separated on 0.5% agarose-formaldehyde gels and transferred to Hybond N+ paper, fixed and (Figure 2A) radioactively probed with HCV specific minus-strand RNA that detects the presence of plus-strand replicon RNA. Lanes 1 and 2: positive controls that contain 109 copies of in vitro transcribed APGK12 RNA. Lane 3: negative control of total cellular RNA from untransfected 20 Huh-7 cells. Lanes 4 and 5: cellular RNA from B1 and B3 cell lines that have integrated DNA copies of the neomycin phosphotransferase gene. Lane 6: total cellular RNA from a Huh-7 cell line, designated S22.3, that harbors high copy number HCV sub-genomic replicon RNA as highlighted by the arrow. Other cell lines 25 have no detectable replicon RNA. Figure 2B is identical to Figure 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA. Lanes 1 and 2 are positive control lanes that contain 109 copies of full length HCV minus strand RNA. Lane 6, which contains 12 µg of total cellular RNA from cell line S22.3, harbors detectable minusstrand replicon RNA at the expected size of 8 - 9 kilobases. M represent the 30 migration of non-radioactive molecular size markers on the agarose gel. 28s represents the migration of 28s ribosomal RNA and accounts for the detection of this species in a samples of total cellular RNA.
- 35 Figure 3 shows indirect immunofluorescence of a HCV non-structural protein in the

8

S22.3 cell line. Indirect immunofluorescence was performed on cells that were cultured and fixed, permeabilized and exposed to a rabbit polyclonal antibody specific for a segment of the HCV NS4A protein. Secondary goat anti-rabbit antibody conjugated with red-fluor Alexa 594 (Molecular Probes) was used for detection. Top panels shows the results of immunofluorescence (40X objective) and the specific staining of the S22.3 cells. The bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCM NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have background level of staining.

5

10

15

Figure 4 shows Western-blots following SDS-PAGE separation of total proteins extracted from three cell lines: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary \ goat anti-rabbit antibody. Panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than control B1 cells and that the naïve Huh-7 cell line does not produce the NPT protein. Panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. M represents molecular weight (in kilodaltons) of pre-stained polypeptide markers.

Figure 5A and 5B identify the nucleotide and amino acid sequences respectively that differ from the APGK12 sequence in the different HCV bi-cistronic replicons. The S22.3 adapted replicon is a first generation replicon selected following the transfection of RNA transcribed from the APGK12 template. R3, R7, R16 are second generation replicons that were selected following the transfection of RNA isolated from the S22.3 first generation replicon cell line. Figure 5A: Nucleotide mutations that were characterized in each of the adapted replicons are indicated adjacent to the respective segment of the replicon (IRES, NS3, NS4A, NS5A, and NS5B). Figure 5B: Amino acid numbers are numbered according to the full length HCV poly-protein with the first amino acid in the second cistron corresponding to amino acid 810 in NS2 of I377/NS2-3' construct.

9

Figure 6 depicts the colony formation efficiency of four in vitro transcribed HCV subgenomic bi-cistronic replicon RNAs. The APGK12 serves as the reference sequence; highlighted are the initiating nucleotides of the HCV IRES in each of the constructs and the amino acid differences (from the APGK12 reference sequence) in the HCV non-structural region for the two R3-rep. Note that the in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in panels A and B) following selection with 0.25 mg/ml G418. RNA isolated from the second generation R3 cell line was reverse transcribed into DNA and cloned into the pAPGK12 vector backbone to generate the R3-rep, which was sequenced and found to encode additional changes that included the L(2155)P substitution in the NS5A segment of the HCV polyprotein (compare R3-rep sequence with the R3 sequence in tables 2 and 3). Various quantities of in vitro transcribed R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10⁶ cfu/µg of RNA (panel C). Various quantities of R3-rep-5'G were also transfected resulting in a colony formation efficiency of 2 X 10⁶ cfu/µg of RNA (panel D).

Figure 7 displays a typical RT-PCR amplification plot (left panel) and the graphical representation of Ct values versus known HCV RNA quantity in a standard curve (right panel). Each of the plotted curves in the left panel, graph the increment of fluorescence reporter signal (delta-Rn) versus PCR cycle number for a predetermined quantity of HCV replicon RNA. The Ct value is obtained by determining the point at which the fluorescence exceeds an arbitrary value (horizontal line). The right panel demonstrates the linear relationship between starting RNA copy number of the predetermined standards (large black dots) and the Ct value. Smaller dots are the Ct values of RNA samples (containing unknown quantity of HCV replicon RNA) from S22.3 cells treated with various concentrations of a specific inhibitor of HCV replication.

30

35

5

10

15

20

25

Figure 8 shows the effect of increasing concentration of inhibitor A on HCV RNA replicon levels in Huh7 cells. S22.3 cells were grown in the presence of increasing concentrations of inhibitor A starting at 0.5nM and ranging to 1024nM. The inhibitor dose-response curve is the result of 11 concentrations from serial two-fold dilutions (1:1). One control well, without any inhibitor, was also included during the course of

the experiment. The cells were incubated for 4 days in a 5% CO₂ incubator at 37 °C. Total cellular RNA was extracted, quantified by optical density. HCV replicon RNA was evaluated by real time RT-PCR and plotted as genome equivalents/µg total RNA as a function of inhibitor concentration

5

10

15

Definitions

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell culture, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook *et al.* (1989) and Ausubel *et al.* (1994).

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction,

from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission (1972).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

- 20 The term "DNA segment or molecule or sequence", is used herein, to refer to molecules comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). These segments, molecules or sequences can be found in nature or synthetically derived. When read in accordance with the genetic code, these sequences can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.
 - As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.
- A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific structural function that constitute the viral particles. "Structural proteins" defines the HCV proteins incorporated into the virus particles namely, core "C", E1, E2, and E2-p7.
- "Non-structural proteins", defines the HCV proteins that are not comprised in viral particles namely, NS2, NS3, NS4A, NS5A and NS5B.

11

"Restriction endonuclease or restriction enzyme" is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5 or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. An example of such an enzyme is *EcoRI*, which recognizes the base sequence G↓AATTC and cleaves a DNA molecule at this recognition site. "Restriction fragments" are DNA molecules produced by the digestion of DNA with a restriction endonuclease. Any given genome or DNA segment can be digested by a particular restriction endonuclease into at least two discrete molecules of restriction fragments.

5

25

35

"Agarose gel electrophoresis" is an analytical method for fractionating polynucleotide molecules based on their size. The method is based on the fact that nucleic acid molecules migrate through a gel as through a sieve, whereby the smallest molecule has the greatest mobility and travels the farthest through the gel. The sieving characteristics of the gel retards the largest molecules such that, these have the least mobility. The fractionated polynucleotides can be visualized by staining the gel using methods well known in the art, nucleic acid hybridization or by tagging the fractionated molecules with a detectable label. All these methods are well known in the art, specific methods can be found in Ausubel et al. (supra).
"Oligonucleotide or oligomer" is a molecule comprised of two or more

"Oligonucleotide or oligomer" is a molecule comprised of two or more

deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size
of the molecule will depend on many factors, which in turn depend on the ultimate
function or use of the oligonucleotide. An oligonucleotide can be derived
synthetically, by cloning or by amplification.

"Sequence amplification" is a method for generating large amounts of a target sequence. In general, one or more amplification primers are annealed to a nucleic acid sequence. Using appropriate enzymes, sequences found adjacent to, or in between the primers are amplified. An amplification method used herein is the polymerase chain reaction (PCR) and can be used in conjunction with the reverse-transcriptase (RT) to produce amplified DNA copies of specific RNA sequences.

"Amplification primer" refers to an oligonucleotide, capable of annealing to a RNA or DNA region adjacent to a target sequence and serving as the initiation primer for DNA synthesis under suitable conditions well known in the art. The synthesized primer extension product is complementary to the target sequence.

The term "domain" or "region" refers to a specific amino acid sequence that defines either a specific function or structure within a protein. As an example herein, is the

5

10

· 15

20

25

30

35

NS3 protease domain comprised within the HCV non-structural polyprotein. The terms "plasmid" "vector" or "DNA construct" are commonly known in the art and refer to any genetic element, including, but not limited to, plasmid DNA, phage DNA, viral DNA and the like which can incorporate the oligonucleotide sequences, or sequences of the present invention and serve as DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The terminology "expression vector" defines a vector as described above but designed to enable the expression of an inserted sequence following transformation or transfection into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. Such expression control sequences will vary depending on whether the vector is designed to express the operably linked gene *in vitro* or *in vivo* in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) or RNA, when such nucleic acid has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting/transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, an example of a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome and is inherited by daughter cells through chromosome replication. A host cell or indicator cell can be transfected with RNA. A cell can be stably transfected with RNA if the RNA replicates and copies of the RNA segregate to daughter cells upon cell division. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA or RNA. Transfection methods are well known in the art (Sambrook et al., 1989; Ausubel et al., 1994). If the RNA encodes for a genetic marker that imparts an observable phenotype, such as antibiotic resistance, then the stable transfection of replicating RNA can be monitored by the acquisition of such phenotype by the host cell.

As used herein the term "transduction" refers to the transfer of a genetic marker to host cells by the stable transfection of a replicating RNA.

WO 02/052015

10

15

20

13

PCT/CA01/01843

The nucleotide sequences and polypeptides useful to practice the invention include without being limited thereto, mutants, homologs, subtypes, quasi-species, alleles, and the like. It is understood that generally, the sequences of the present invention encode a polyprotein. It will be clear to a person skilled in the art that the polyprotein of the present invention and any variant, derivative or fragment thereof, is auto-processed to an active protease.

As used herein, the designation "variant" denotes in the context of this invention a sequence whether a nucleic acid or amino acid, a molecule that retains a biological activity (either functional or structural) that is substantially similar to that of the original sequence. This variant may be from the same or different species and may be a natural variant or be prepared synthetically. Such variants include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided the biological activity of the protein is conserved. The same applies to variants of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained.

The term "derivative" is intended to include any of the above described variants when comprising additional chemical moiety not normally a part of these molecules. These chemical moieties can have varying purposes including, improving a molecule's solubility, absorption, biological half life, decreasing toxicity and eliminating or decreasing undesirable side effects. Furthermore, these moieties can be used for the purpose of labeling, binding, or they may be comprised in fusion product(s). Different moieties capable of mediating the above described effects can be found in *Remington's The Science and Practice of Pharmacy* (1995).

Methodologies for coupling such moieties to a molecule are well known in the art.

The term "fragment" refers to any segment of an identified DNA, RNA or amino acid sequence and/or any segment of any of the variants or derivatives described herein above that substantially retains its biological activity (functional or structural) as required by the present invention.

The terms "variant", "derivative", and "fragment" of the present invention refer herein to proteins or nucleic acid molecules which can be isolated/purified, synthesized chemically or produced through recombinant DNA technology. All these methods are well known in the art. As exemplified herein below, the nucleotide sequences and polypeptides used in the present invention can be modified, for example by *in vitro* mutagenesis.

:

10

20

35

As used herein, the term "HCV polyprotein coding region" means the portion of a hepatitis C virus that codes for the polyprotein open reading frame (ORF). This ORF may encode proteins that are the same or different than wild-type HCV proteins. The ORF may also encode only some of the functional protein encoded by wild-type polyprotein coding region. The protein encoded therein may also be from different isolates of HCV, and non-HCV protein may also be encoded therein.

As used herein, the abbreviation "NTR" used in the context of a polynucleotide molecule means a non-translated region. The term "UTR" means untranslated region. Both are used interchangeably.

Preferred embodiments

Particularly, the invention provides a HCV self-replicating polynucleotide molecule comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO.8).

According to the first embodiment of this invention, there is particularly provided a HCV polynucleotide construct comprising:

- a 5'-non translated region (NTR) comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein coding region; and
- a 3'-NTR region.

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

Alternatively, the first embodiment of the present invention is directed to HCV self-replicating polynucleotide molecule comprising a G2042C/R mutation.

5

According to the second embodiment, the present invention particularly provides a HCV polynucleotide construct comprising:

- a 5'-NTR region comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein region coding for a HCV polyprotein comprising a
 G(2042)C or a G(2042)R mutation; and
- a 3'-NTR region.
- Preferably, the polynucleotide construct of the present invention is a DNA or RNA molecule. More preferably, the construct is a RNA molecule. Most preferably, the construct is a DNA molecule.
- More particularly, the first embodiment of this invention is directed to a RNA

 molecule encoded by the DNA molecule selected from the group consisting of: SEQ

 ID NO. 2, 4, 5, 6, 7, 24 and 25.
 - Most particularly, the invention provides a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.

In a third embodiment, the invention also is directed to an expression vector comprising DNA forms of the above polynucleotide, operably linked with a promoter.

Preferably, the promoter is selected from the group consisting of: T3, T7 and SP6.

25

20

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or vector as described above. Particularly, the host cell is a eukaryotic cell line. More particularly, the eukaryotic cell line is a hepatic cell line. Most particularly, the hepatic cell line is Huh-7.

30

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- a) incubating the host cell as described above under conditions suitable for RNA replication;
- b) isolating the total cellular RNA from the cells; and

c) analyzing the RNA so as to measure the amount of HCV RNA replicated.

Preferably, the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

Alternatively in this fifth embodiment, the construct comprises a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.

10

15 .

5

According to a preferred aspect of the sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- a) carrying step a) as described in the above assay, in the presence or absence of the compound;
- b) isolating the total cellular RNA from the cells; and
- c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,
- wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

Preferably, the cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

25

EXAMPLES

EXAMPLE 1

Replicon Constructs (APGK-12; Figure 1)

pET9a-EMCV was obtained by ligating an oligonucleotide linker
5' gaattccagatggcgcccagatgttaaccagatccatggcacactctagagtactgtcgac 3' (SEQ ID NO.9) to pET-9a (Novagen) that was cut with EcoRI and Sall to form the vector pET-9a-mod. This linker contains the following restriction sites: EcoRI, AscI, HpaI, NcoI, XbaI, ScaI, Sall. The EMCV IRES was amplified by PCR from the vector pTM1 with primers

17

5' cggaatcgttaacagaccacaacggtttccctc 3' (SEQ ID NO.10) and 5' ggcgtacccatggtattatcgtgtttttca 3' (SEQ ID NO.11) and ligated into pET-9a-mod via EcoRI and NcoI to form pET-9a-EMCV.

- The sequence of HCV NS2 to NS5B followed by the 3'UTR of HCV was obtained from the replicon construct I377/NS2-3' (Lohman et al., 1999; accession number: AJ242651) and synthesized by Operon Technologies Inc. with a T to C change at the Ncol site in NS5B at nucleotide 8032. This sequence was released from an GenOp® vector (Operon Technologies) with Ncol and Scal and transferred into pET-9a-EMCV-NS2-5B-3'UTR.
 - pET-9a-HCV-neo was obtained by amplification of the HCV IRES from a HCV cDNA isolated from patient serum with primers
- 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.12) containing a
 T7 promoter and primer
 - 5' ggcgcgccctttggtttttctttgaggtttaggattcgtgctcat 3' (SEQ ID NO.13) and amplification of the neomycin phosphotransferase gene from the vector pcDNA 3.1 (Invitrogen) with primers
 - 5' aaagggcgcatgattgaacaagatggattgcacgca 3' (SEQ ID NO.14) and 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15). These two PCR fragments were mixed and amplified with primers
 - 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.16) and 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15); cut with Eco RI and HpaI and transferred into pET-9a-mod to form pet-9a-HCV-neo. The EMCV-
- NS2-5B-3'UTR was released from pET-9a-EMCV-NS2-5B-3'UTR with HpaI and ScaI and transferred into pet-9a-HCV-neo that was cut with HpaI to form pET-9a-APGK12. This insert was sequenced with specific successive primers using a ABI Prism® BigDye™ Terminator Cycle sequencing kit and analyzed on ABI Prism® 377 DNA Sequencer and is shown in SEQ ID NO 1.

30

35

20

RNA in vitro transcription

pET-9a-APGK12 DNA was cut with Scal for expression of the full-length replicon or with BgllI for expression of a truncated negative control RNA. DNA was analyzed on a 1% agarose gel and purified by Phenol/Chloroform extraction. RNA was produced using a T7 Ribomax® kit (Promega) followed by extraction with phenol/chloroform

18

and precipitation with 7.5 M LiCl₂. RNA was treated with DNAse I for 15 min to remove the DNA template and further purified with an RNeasy® column (Qiagen). RNA integrity was verified on a denaturing formaldehyde 1% agarose gel.

5 EXAMPLE 2

10

15

20

25

30

Primary transfection of Huh7 cells and selection of replicon cell lines
Human hepatoma Huh7 cells (Health Science Research Resources Bank, Osaka, Japan) were grown in 10% FBS/DMEM. Cells were grown to 70% confluency, trypsinized, washed with phosphate buffered saline (PBS) and adjusted to 1x10⁷ cells/ml of PBS. 800 μl of cells were transferred into 0.4cm cuvettes and mixed with 15 μg of replicon RNA. Cells were electroporated using 960μF, 300 volts for ~18 msec and evenly distributed into two 15 cm tissue culture plates and incubated in a tissue culture incubator for 24 hours. The selection of first and second generation replicon cell lines was with 10% FBS/DMEM medium supplemented with 1mg/ml of G418. Cells were selected for 3-5 weeks until colonies were observed that were isolated and expanded.

Following the G418 selection and propagation of Huh-7 cells transfected with APGK12 (SEQ ID NO. 1) RNA, cells that formed a distinct colony were treated with trypsin and serially passed into larger culture flasks to establish cell lines. Approximately 10 X 10⁶ cells were harvested from each cell line. The cells were lysed and the total cellular RNA extracted and purified as outlined in Qiagen RNAeasy® preparatory procedures. Figure 2 shows the analysis of 12 µg of total cellular RNA from various cell lines as analyzed on a Northern blot of a denaturing agarose-formaldehyde gel.

Figure 2A is a Northern blot (radioactively probed with HCV specific minus-strand RNA) that detects the presence of plus-strand replicon RNA. Lanes 1 and 2 are positive controls that contain 10⁹ copies of in vitro transcribed APGK12 RNA. Lane 2 contains the *in vitro* transcribed RNA mixed with 12 µg of total cellular from naïve Huh-7 cells. Lane 3 is a negative control of total cellular RNA from untreated Huh-7 cells. Lanes 4 and 5 contain cellular RNA from the B1 and B3 G418 resistant cell lines that have DNA integrated copies of the neomycin phosphotransferase gene. Lane 6 contains total cellular RNA from a Huh-7 cell line, designated S22.3, that

19

harbors high copy number of HCV sub-genomic replicon RNA as detected by the positive signal in the 8 kilo-base range. Other cell lines have no detectable replicon RNA. Figure 2B is a Northern blot of a duplicate of the gel presented in 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA (lanes 1 and 2 are positive control lanes that contain 10^9 copies of full length genomic HCV minus strand RNA); only lane 6, which contains $12~\mu g$ of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8-9 kilobases. An quantitative estimation of RNA copy number, based on phosphorimager scanning of the Northern blots, is approximately $6~X10^7$ copies of plus-strand/ μg of total RNA, and $6~x~10^6$ copies of minus strand/ μg of total RNA. The presence of the plus-strand and minus-strand intermediate confirms that the HCV sub-genomic RNA is actively replicating in the S22.3 cell line.

15 Example 3

5

10

20

25

30

35

S22.3 cell line constitutively expresses HCV non-structural proteins.

HCV non-structural protein expression was examined in the S22.3 cell line. Figure 3 displays the result of indirect immunofluorescence that detects the HCV NS4A protein in the S22.3 cell line and not in the replicon negative B1 cell line (a G418 resistant Huh-7 cell line). Indirect immunofluorescence was performed on cells that were cultured and fixed (with 4% paraformaldehyde) onto Lab-tek chamber slides. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes followed by a 1 hour treatment with 5% milk powder dissolved in phosphate-buffered saline (PBS), A rabbit serum containing polyclonal antibody raised against a peptide spanning the HCV NS4A region was the primary antibody used in detection. Following a 2 hour incubation with the primary antibody, cells were washed with PBS and a secondary goat anti-rabbit antibody conjugated with red-fluor Alexa® 594 (Molecular Probes) was added to cells for 3 hours. Unbound secondary antibody was removed with PBS washes and cells were sealed with a cover slip. Figure 3 (top panels) shows the results of immunofluorescence as detected by a microscope with specific fluorescent filtering; the bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have

background level of staining. A small proportion of S22.3 cells express high levels of intensely stained HCV NS4A.

Expression of the proteins encoded by the bi-cistronic replicon RNA was also examined on Western-blots following SDS-PAGE separation of total proteins extracted from: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Figure 4 panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary HRP-conjugated goat anti-rabbit antibody. Figure 4 panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than B1 cells (which contain an integrated DNA copy of the *npt* gene) and that the naïve Huh-7 cell line does not produce the NPT protein. Figure 4 panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. The western blots demonstrate that the S22.3 cell line, which harbors actively replicating HCV sub-genomic replicon RNA, maintains replication of the RNA through the high level expression of the HCV non-structural proteins.

20 EXAMPLE 4

5

10

15

25

30

Sequence determination of adapted replicons

Total RNA was extracted from replicon containing Huh7 cells using a RNeasy Kit (Qiagen). Replicon RNA was reverse transcribed and amplified by PCR using a OneStep RT-PCR kit (Qiagen) and HCV specific primers (as selected from the full-length sequence disclosed in WO 00/66623). Ten distinct RT-PCR products, that covered the entire bi-cistronic replicon in a staggered fashion, were amplified using oligonucleotide primers. The PCR fragments were sequenced directly with ABI Prism® BigDye™ Terminator Cycle PCR Sequencing and analyzed on ABI Prism® 377 DNA Sequencer. To analyze the sequence of the HCV replicon 3' and 5' ends a RNA ligation/RT-PCR procedure described in Kolykhalov *et al.* 1996 was followed. The nucleotide sequence of S22.3 is presented as SEQ ID NO. 2.

21

EXAMPLE 5.

Serial Passage of HCV Replicon RNA

The total cellular RNA from the S22.3 cell line was prepared as described above. HCV Replicon RNA copy number was determined by Taqman® RT-PCR analysis and 20 μg of total S22.3 cellular RNA (containing 1 X 109 copies of HCV RNA) was transfected by electroporation into 8 X 10⁶ naïve Huh-7 cells. Transfected cells were subsequently cultured in 10 cm tissue culture plates containing DMEM supplemented with 10% fetal calf serum (10% FCS). Media was changed to DMEM (10% FCS) supplemented with 1 mg/ml G418 24 hours after transfection and then 10 changed every three days. Twenty-three visible colonies formed three to four weeks post-transfection and G418 selection. G418 resistant colonies were expanded into second generation cell lines that represent the first cell lines harboring serially passaged HCV Replicon RNA. Three of these cell lines: R3, R7, and R16 were the subject of further analyses. First, the efficiency of transduction by each of the 15 adapted replicons was determined by electroporation of the total cellular RNA (extracted from the R3, R7 and R16) into naïve Huh-7 cells; following electroporation, the transduction efficiency was determined as described above, by counting the visible G418 resistant colonies that arose following 3 to 5 weeks of 20 ¹ G418 selection (Table 1). Second, the sequence of the serially passed adapted replicons was determined from the total cellular RNA that was extracted from each of the R3, R7 and R16 replicon cell lines as described in example 4 (SEQ ID NO. 4, 5, 6). Using the pAPGK12 as a reference sequence (SEQ ID NO. 1), the nucleotide changes that were selected in HCV segment of the adapted replicons are presented 25 in Figure 5A. Some of these nucleotide changes are silent and do not change the encoded amino acid whereas others result in an amino acid substitution. Figure 5B summarizes the amino acid changes encoded by the adapted replicons with the amino acid sequence of pAPGK12 as the reference. It is important to note that the reference sequence APGK-12 (SEQ ID NO.1) contains an extra G at the 5'-terminal 30 (5'-GG) that is not maintained in the replicating RNA of the established cell lines. Also noteworthy is that, in addition to G->A at nucleotide 1, there is also an adapted mutation G->C/R at amino acid 2042 (shown as amino acid 1233 in the sequence listing since a.a. 810 of NS2 is numbered as a.a. 1 in SEQ ID) that can be found in all clones analyzed.

22

TABLE 1
Transfection of Huh-7 cells

	<u>RNA</u>	Copies of Replicon	# Colonies	SEQ ID
5				
	5 ng APKG12 replicon in 20μg total Huh-7 RNA	1.2 x 10 ⁹	0	
10	15 μg APKG12 replicon RNA	3 x 10 ¹²	1 (S22.3)	1
	20μg total:			
	S22.3 cellular RNA	3 x 10 ⁹	23 (3 clones analyzed)	2
15	R3 cellular RNA	1 x 10 ⁹	200	4
	R7 cellular RNA	1 x 10 ⁹	20	5
	R16 cellular RNA	3 x 10 ⁸	100	6
	cloned R3rep RNA	<u>2.3 x 10⁸</u>	2000	7

20 EXAMPLE 6

25

Construction of APGK12 with 5' G-> A substitution (APGK12-5'A, SEQ ID NO.24)

The pAPGK12 DNA was modified to change the first nucleotide in the sequence to replace the 5'GG with a 5'A. The change in the pAPGK12 was introduced by replacing an *EcoRl/Age*l portion of the sequence with a PCR-generated *EcoRl/Age*l fragment that includes the mutation. The oligonucleotides used for the amplification were (SEQ ID. NO. 20): 5'-GTG GAC GAA TTC TAA TAC GAC TCA CTA TAA CCA GCC CCC GAT TGG-3' and (SEQ ID. NO. 21): 5'-GGA ACG CCC GTC GTG GCC AGC CAC GAT-3' and generated a 195 bp DNA fragment that was then digested with *EcoRl* and *Agel*. The resulting 178 bp restriction fragment was used to replace the *EcoRl / Agel* fragment in pAPGK12 to generate the pAPGK12-5'A plasmid.

EXAMPLE 7

cDNA CLONING OF THE R3-REPLICON (R3REP).

The cDNA clone of the R3 replicon was produced by RT-PCR of RNA extracted from the R3 cell line. The following two oligonucleotides were used: (SEQ ID. NO. 22): 5'-GTC GTC TTC TCT GAC ATG GAG AC-3' and (SEQ ID. NO. 23): 5'-GAG TTG

CTC AGT GGA TTG ATG GGC AGC-3'. The ~4400nt PCR fragment, starting within the NS2 coding region and extending to the 5'-end of the NS5B coding region, was cloned into the plasmid pCR3.1 by TA cloning (Invitrogen). The SacII / XhoI portion of this R3 sequence was then used to replace the SacII / XhoI fragment present in the pAPGK12 and the pAPGK12-5'A described above. Consequently, two R3 cDNA sequences were generated: (I) R3-Rep-5'G with an initiating 5'G (SEQ ID NO.7), and R3-Rep-5'A (SEQ ID NO.25) with an initiating 5'A. Sequencing of the R3 rep cDNA identified unique nucleotide changes that differ from the original pAPGK12 sequence (see Figure 5A); some of these changes are silent and do not change the encoded amino acid, whereas others do result in an amino acid change (see Figure 5B). The differences between R3 and the R3-rep reflect the isolation of a unique R3-rep cDNA clone encoding nucleotide changes that were not observed from the sequencing of the total RNA extracted from the R3 cell line.

15 EXAMPLE 8

5

10

20

25

30

35

Efficiency of colony formation with modified constructs

RNA from pAPGK12, pAPGK12-5'A, pR3-Rep and pR3-Rep-5'A was generated by in vitro transcription using the T7 Ribomax® kit (Promega) as described in example 1 above. The reactions containing the pAPGK12-5'A and pR3-Rep-5'A templates were scaled-up 10-fold due to the limitation of commercial RNA polymerase in initiating transcripts with 5'-A. The full length RNAs and control truncated RNA for each clone were introduced into 8 x 10⁶ naïve Huh-7 cells by electroporation as described in example 2. Replicon RNA was supplemented with total cellular Huh-7 carrier RNA to achieve a final 15-20µg quantity. The cells were then cultured in DMEM medium supplemented with 10% fetal calf serum and 0.25 mg/ml G418 in two 150 mm plates. The lower concentration of G418 was sufficient to isolate and select replicon containing cell lines as none of the transfectants with the control truncated RNA produced any resistant colonies. In contrast, in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in Figure 6 panels A and B) following selection with G418. Various quantities (ranging from 0.1 ng to 1 µg) of the R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10⁶ cfu/µg of RNA (Figure 6 panel C depicts transfection with 1 µg of RNA). Various quantities (ranging from 0.1 ng to 1 µg) of R3-rep [5'G] were similarly transfected resulting in a colony formation efficiency of 2 X 10⁶ cfu/ug of RNA (Figure 6 panel D

24

depicts colony formation with 1µg of RNA). Note that, shown for the first time, HCV subgenomic replicons replicate as efficiently with a 5' A nucleotide in place of the 5'G. APGK12 with a 5'A or 5'G RNA have similar transduction efficiencies. Similarly, R3-Rep RNAs with either the 5'A or 5'G both display the markedly increased transduction efficiency. Notably, the adaptive mutants within the HCV non-structural segment encoded by the R3-Rep provides for a substantial increase in transduction efficiency as depicted by the dramatic increase in colony forming units per µg of transfected RNA.

10 EXAMPLE 9

5

15

20

25

Quantification of HCV Replicon RNA Levels in Cell lines

S22.3 cells, or cell lines harboring other adapted replicons, were seeded in DMEM supplemented with 10% FBS, PenStrep and 1μg/mL Geneticin. At the end of the incubation period the replicon copy number is evaluated by real-time RT-PCR with the ABI Prism 7700 Sequence Detection System. The TAQMAN® EZ RT-PCR kit provides a system for the detection and analysis of HCV RNA (as first demonstrated by Martell *et al.* 1999 J. Clin. Microbiol. 37: 327-332). Direct detection of the reverse transcription polymerase chain reaction (RT-PCR) product with no downstream processing is accomplished by monitoring the increase in fluorescence of a dyelabeled DNA probe (Figure 6). The nucleotide sequence of both primers (adapted from Ruster, B. Zeuzem, S. and Roth, W.K., 1995. Analytical Biochemistry 224:597-600) and probe (adapted from Hohne, M., Roeske, H. and Schreier, E. 1998, Poster Presentation: P297 at the Fifth International Meeting on Hepatitis C Virus and Related Viruses Molecular Virology and Pathogenesis, Venezia-Lido Italy, June 25-28, 1998) located in the 5'-region of the HCV genome are the following:

HCV Forward primer:

5' ACG CAG AAA GCG TCT AGC CAT GGC GTT AGT 3' (SEQ ID NO.17)

30

HCV Reverse primer:

5' TCC CGG GGC ACT CGC AAG CAC CCT ATC AGG 3' (SEQ ID NO.18)

HCV Probe:

5' FAM-TGG TCT GCG GAA CGG GTG AGT ACA CC-TAMRA 3' (SEQ ID NO.19)

5 FAM: Fluorescence reporter dye.

TAMRA: Quencher dye.

Using The TAQMAN® EZ RT-PCR kit, the following reaction was set up:

Component	Volume per sample	Final	
	(µL)	Concentration	
RNase-Free Water	16	-	
5X Taqman EZ Buffer	10	1X	
Manganese Acetate 25mM	6	3mM	
dATP 10mM	1.5	300µM	
dCTP 10mM	1.5	300µM	
dGTP 10mM	1.5	300µM	
dUTP 20mM	1.5	300µM	
HCV Forward Primer 10µM	1	200nM	
HCV Reverse Primer 10µM	1	200nM	
HCV Probe 5uM	2	200nM	
rTth DNA Polymerase	2	0.1U/µL	
2.5U/µL			
AmpErase UNG 1U/μL	0.5	0.01U/µL	
Total Mix	45	-	

10

15

To this reaction mix, 5μ L of total RNA extracted from S22.3 cells diluted at $10 \text{ng}/\mu$ L was added, for a total of 50 ng of RNA per reaction. The replicon copy number was evaluated with a standard curve made from known amounts of replicon copies (supplemented with 50 ng of wild type Huh-7 RNA) and assayed in an identical reaction mix (Figure 7).

Thermal cycler parameters used for the RT-PCR reaction on the ABI Prism 7700 Sequence Detection System were optimized for HCV detection:

26

Cycle	Temperature (°C)	Time (Minutes)	Repeat	Reaction
Hold	50	2		Initial Step
Hold	60	30		Reverse
		•		Transcription
Hold	95	5		UNG Deactivation
Cycle	95	0:15	2	Melt
Cycle	60	1	2	Anneal/Extend
Cycle	90	0:15	40	Melt
	60	1	40	Anneal/Extend

Quantification is based on the threshold cycle, where the amplification plot crosses a defined fluorescence threshold. Comparison of the threshold cycles provides a highly sensitive measure of relative template concentration in different samples. Monitoring during early cycles, when PCR fidelity is at its highest, provides precise data for accurate quantification. The relative template concentration can be converted to RNA copy numbers by employing a standard curve of HCV RNA with known copy number (Figure 7).

10 **EXAMPLE 10**

15

20

A specific HCV NS3 protease anti-viral compound inhibits replication of the HCV replicon in S22.3 cell lines.

In order to determine the effect of a specific HCV NS3 protease anti-viral compound on replicon levels in S22.3 cells, the cells were seeded in 24 Well Cell Culture Cluster at 5 X 10⁴ cells per well in 500μL of DMEM complemented with 10% FBS, PenStrep and 1μg/mL Geneticin. Cells were incubated until compound addition in a 5% CO₂ incubator at 37 °C. The dose-response curve of the inhibitor displayed 11 concentrations resulting from serial two-fold dilutions (1:1). The starting concentration of compound A was 100nM. One control well (without any compound) was also included in the course of the experiment. The 24 well plates were incubated for 4 days in a 5% CO₂ incubator at 37 °C. Following a 4 day incubation period, the cells were washed once with PBS and RNA was extracted with the RNeasy® Mini Kit and Qiashredder® from Qiagen. RNA from each well was eluted in 50uL of H₂O. The RNA was quantified by optical density at 260nm on a Cary 1E UV-Visible Spectrophotometer. 50 ng of RNA from each well was used to quantify the HCV replicon RNA copy number as detailed in Example 6. The level of inhibition (% inhibition) of each well containing inhibitor was calculated with the following

27

equation (CN = HCV Replicon copy number):

$$\% \cdot inhibition = \left(\frac{CN \cdot control - CN \cdot well}{CN \cdot control}\right) * 100$$

The calculated % inhibition values were then used to determine IC₅₀, slope factor (n) and maximum inhibition (I_{max}) by the non-linear regression routine NLIN procedure of SAS using the following equation:

$$\% \cdot inhibition = \frac{I_{\text{max}} \times [inhibitor]^{n}}{[inhibitor]^{n} + IC_{50}^{n}}$$

10

Compound A was tested in the assay at least 4 times. The IC_{50} curves were analyzed individually by the SAS nonlinear regression analysis. Figure 8 shows a typical curve and Table 2 shows the individual and average IC_{50} values of compound A. The average IC_{50} of compound A in the replication assay was 1.1nM.

15

TABLE 2

IC₅₀ of compound A in the S22.3 Cell line Replicon Assay.

Compound	IC ₅₀ (nM)	Average IC ₅₀ (nM)
	1.2	•
Α	1.2	
	1.0	
	0.9	
		1.1 + 0.2

 1.1 ± 0.7

20 DISCUSSION

25

The reproducible and robust *ex vivo* propagation of hepatitis C virus, to levels required for the accurate testing of potential anti-viral compounds, has not been achieved with any system. As an alternative approach to studying the molecular mechanisms of hepatitis C virus RNA replication, selectable self-replicating bicistronic RNAs were developed (Lohman *et al.*, 1999, Science 285:110-113; Bartenschlager CA 2,303,526). Minimally, these replicons encode for some or all of

28

the non-structural proteins and also carry a selectable marker such as the neomycin phosphotransferase. Though intracellular steady-state levels of these sub-genomic replicon RNAs among the selected clones is moderate to high, the frequency of generating G418-resistant colonies upon transfection of the consensus RNA described by Lohman et al. or Bartenschlager is very low. Less than 100 colonies are generated when 8 million cells are transfected with 1 µg of in vitro transcribed bicistronic replicon RNA. A low efficiency of colony formation was first noted by Lohmann et al (1999 et al, Science 285:110-113). Since then, Lohmann et al. (2001), Blight et al. (2000), and Guo et al. (2001), have isolated sub-genomic RNAs with markedly improved efficiencies in the colony formation assay. Lohmann et al., 1999 originally reported that selection of sub genomic replicons may not involve the selection of adaptive mutants as serially passaged RNA did not demonstrate an improved transfection efficiency. Nevertheless, in an effort to characterize the function and fitness of replicating HCV RNA, we serially passaged the replicon RNA that was isolated from the first selected cell-line. Notably, a significant increase in colony forming efficiency was obtained from this experiment, even though the quantity of replicon RNA was orders of magnitude lower than originally used to transfect the in vitro transcribed RNA. Furthermore, a second round serial passage of replicon RNA from this first generation clone into naive Huh-7 cells provided for yet another increase in colony formation efficiency (Table 1).

Our analysis of replicating HCV RNAs identified several adaptive mutations that enhance the efficiency of colony formation by up to 4 orders of magnitude. Adaptive mutations were found in many non-structural proteins, as well as in the 5' non-translated region. The substitution of the 5'-GG doublet for a 5'-A as the inaugurating nucleotide of the HCV 5'-UTR is a variant of the HCV genome that has not been previously described, despite the sequencing of innumerable genotypes and subtypes from across the world. Our original replicon that carried a 5'-GG evolved to variants with either a single 5'-A or 5'-G, both of which showed equal transduction efficiency. We describe here the first report of a HCV genome that can tolerate and stably maintain a 5'A extremity. Moreover, we were successful in re-introducing this defined single nucleotide substitution into our cDNA clone and generate *in vitro* transcribed RNA harboring such an extremity to confirm that a 5'A functions as efficiently as a 5'G.

10

15

20

25

30

We have identified adaptive amino acid substitutions in the HCV non-structural proteins NS3, NS4A and NS5A in the R3 replicon, and a substitution in NS5B in the R7 clone (see Figure 5B). These mutations, particularly the combination defined by the R3-rep (SEQ ID NO. 7), when reconstituted into a cDNA clone and transcribed onto a RNA replicon, result in a significantly enhanced transduction efficiency of up to 20,000 fold from the original wild type APGK12 replicon RNA. However, the steady state levels of intracellular replicon RNA were comparable from each of the different isolated clones. This result suggests that the increase in replication efficiency by the adaptive mutations does not result in higher stable intracellular RNA levels due to higher RNA replication, but rather confers increased permissivity for establishing the replicon in a greater number of Huh7 cells. Such a phenotype may be manifested transiently, through an initial increase of the amount of *de novo* replication, that is required to surpass a defined threshold to establish persistently replicating RNAs within a population of dividing cells.

Recently three other groups also identified other distinct adaptive mutants. Lohmann et al. (2000) reported enhanced transduction efficiencies of up to 10,000 fold with mutations in NS3, NS4B, NS5A and NS5B. Blight et al. (2000) reported an augmentation of transduction efficiencies up to 20,000 fold with a single mutation in NS5A whereas Guo et al. (2001) reported increases in transduction efficiencies of 5,000-10,000 fold with a deletion of a single amino acid in NS5A. The amino acid substitutions that we describe here have not previously been identified as adaptive mutants that enhance the efficiency of RNA transfection and/or replication. One exception is the mutation of E1202G in NS3 that we found in both the R7 and R16 replicons. This adaptation was previously described by Guo et al (2001) and Krieger et al (2001). All other adaptive mutations, without exception, described herein are unpublished.

The development of selectable subgenomic HCV replicons has provided for potential avenues of exploration on HCV RNA replication, persistence, and pathogenesis in cultured cells. However, the low transduction efficiency with the HCV RNA-containing replicons as originally described (Lohmann et al., 1999) showed that it was not a practical system for reverse genetics studies. The adaptive mutants described herein overcome the low transduction efficiency. In light of the recent descriptions of adaptive mutants by other groups, we note that adaptation can be

achieved by distinct mutations in different HCV NS proteins, although the level of adaptation can vary drastically. The replicons encoding adaptive mutants that are described herein are ideally suited for reverse genetic studies to identify novel HCV targets or host cell targets that may modulate HCV RNA replication or HCV replicon RNA colony formation. The adapted and highly efficient replicons are suitable tools for characterizing subtle genotypic or phenotypic changes that affect an easily quantifiable transduction efficiency.

Lastly, we have used our adapted HCV sub genomic replicon cell-line to

demonstrate the proficient inhibition of HCV RNA replication by a specific small molecule inhibitor of the HCV NS3 protease. This is the first demonstration that an antiviral, designed to specifically inhibit one of the HCV non-structural proteins, inhibits HCV RNA replication in cell culture. Moreover, this compound and our S22.3 cell line validate the proposal that RNA replication is directed by the HCV non-structural proteins NS3 to NS5B. The assay that we have described and validated will be extremely useful in characterizing other inhibitors of HCV non-structural protein function in cell culture in a high throughput fashion.

All references found throughout the present disclosure are herein incorporated by reference whether they be found in the following list or not.

References

Ago et al. 1999, Structure 7: 1417-1426

Ausubel et al., 1994, Current Protocols in Molecular Biology, Wiley, New York.

25 Bartenschlager, R. et al., 1993, J. Virol., 67, 3835-3844.

Behrens et al., 1998, J. Virol. 72, 2364

Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039

Blight et al. 2000, Science 290: 1972-1974

Cho et al., 1998, J. Biol. Chem., 273, 15045

30 Dash et al., 1997, Am. J. Pathol., 151, 363 – 373

Fourner et al. 1998, J. Gen. Virol. 79, 2376

Gale Jr. et al. 1997 Virology 230, 217

Grakoui, A. et al., 1993(a), J. Virol. 67, 1385-1395.

Grakoui A, et al., 1993(b), Proc Natl Acad Sci USA, 90, 10583-7

35 Guo et al. (2001) J. Virol. 8516-8523

Hijikata, M. et al., 1991, Proc. Natl. Acad. Sci. USA. 88, 5547-5551.

Hijikata, M. et al., 1993, J. Virol. 67, 4665-4675.

Ikda et al. 1998, Virus Res. 56, 157

Ito et al. 1996, J. Gen. Virol. 77, 1043 - 1054

5 IUPAC-IUB Biochemical Nomenclature Commission, **1972**, Biochemistry, *11*, 1726-1732.

Kolykhalov et al. 1996 J. of Virology, 7, p. 3363-3371

Khromykh et al., 1997, J. Virol. 71, 1497

Kim, D.W. et al., 1995, Biochem. Biophys. Res. Comm., 215, 160-166.

10 Kim et al., 1996, Cell, 87, 343;

Kim et al., 1998, Structure, 6, 89

Kim et al., 1999, Arch. Virol, 144, 329-343.

Krieger et al. (2001) J. Virol. 4614-4624

Kwong AD. et al., 1998, Antiviral Res., 40, 1-18

15 Lanford et al. 1994, Virology 202, 606

Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943

Lohman et al. 1999, Science 285: 110-113

Lohman et al. (2001) J. Virol. 1437-1449

Love, R.A. et al., 1996, Cell, 87, 331-342

20 Martell et al. 1999 J. Clin. Microbiol. 37: 327-332

Mizutani et al. 1996, J. Virol. 70, 7219 – 7223

Moser et al., 1998, J. Virol. 72, 5318

Reed et al., 1997, J. Virol. 71, 7187

Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, Cold Spring

25 Harbor Labs.

Shimizu et al. 1993, PNAS, USA, 90, 6037 – 6041

Yanagi et al., 1999, Proc. Natl. Acad. Sci. USA, 96, 2291-95

Yao et al., 1997, Nature Structural Biology, 4, 463

Yem et al., 1998, Protein Science, 7, 837

30 Yoo et al. 1995, J. Virol., 69, 32 – 38

32

CLAIMS

- 1. A HCV polynucleotide molecule comprising a 5'-non translated region (NTR) wherein guanine at position 1 is substituted for adenine.
- 2. A HCV self-replicating polynucleotide comprising:
 - a 5'-NTR consisting of ACCAGC (SEQ ID NO. 8);
 - a HCV polyprotein region coding for a HCV polyprotein; and
 - a 3'-NTR region.
- The HCV polynucleotide according to claim 2, wherein said polyprotein comprises one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.
- 4. The HCV polynucleotide encoding a polyprotein comprising one or more of the amino acid substitution as defined in claim 3, and further comprising the amino acid substitution E(1202)G.
- The HCV polynucleotide according to claim 3, wherein said substitution is a G2042C or a G2042R mutation.
- 6. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: K(1691)R; and G(2042)C.
- 7. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 8. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 9. The HCV polynucleotide according to claim 3, wherein said substitution is selected

from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.

- 10. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- The HCV polynucleotide according to claim 2, wherein said polynucleotide is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **12.** The HCV polynucleotide according to claim 2, wherein said polynucleotide is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **13.** An expression vector comprising a DNA form of the polynucleotide according to claim 2, operably linked to a promoter.
- **14.** A host cell transfected with the self-replicating polynucleotide molecule according to claim 2.
- 15. A host cell according to claim 14, wherein the host cell is a eukaryotic cell line.
- 16. A host cell according to claim 15, wherein said eukaryotic cell line is a hepatic cell line.
- 17. A host cell according to claim 16, wherein said hepatic cell line is Huh-7.
- **18.** A RNA replication assay comprising the steps of:
 - a) incubating the host cell according to claim 14 under conditions suitable for RNA replication;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 19. The assay according to claim 18, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

- 20. The assay according to claim 18, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- **21.** A method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) carrying step a) according to claim 18, in the presence or absence of the compound;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
 - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

- 22. The method according to claim 21, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.
- 23. A HCV polynucleotide molecule comprising:
 - a 5'-NTR region;
 - a HCV polyprotein region coding for a HCV polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T; and a 3'-NTR region.
- 24. The HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as defined in claim 24, further comprising the amino acid substitution E(1202)G.
- **25.** The polynucleotide according to claim 24, wherein said substitution is a G2042C or a G2042R mutation.
- 26. The HCV polynucleotide according to claim 24, wherein said substitution is selected

- from the group consisting of: K(1691)R; and G(2042)C.
- 27. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 28. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 29. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- 30. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 31. The HCV polynucleotide according to claim 24, wherein said molecule is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 32. The HCV polynucleotide according to claim 24, wherein said molecule is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- **33.** An expression vector comprising a DNA form of the polynucleotide according to claim 24, operably linked to a promoter.
- 34. A host cell transfected with the self-replicating polynucleotide according to claim 24.
- 35. A host cell according to claim 34, wherein the host cell is a eukaryotic cell line.
- **36.** A host cell according to claim 35, wherein said eukaryotic cell line is a hepatic cell line.
- 37. A host cell according to claim 36, wherein said hepatic cell line is Huh-7.

38. A RNA replication assay comprising the steps of:

incubating the host cell according to claim 34 under conditions suitable for RNA replication;

isolating the total cellular RNA from the cells; and analyzing the RNA so as to measure the amount of HCV RNA replicated.

- 39. The assay according to claim 38, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.
- **40.** The assay according to claim 38, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- **41.** A method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) carrying step a) according to claim 38, in the presence or absence of the compound;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
 - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,

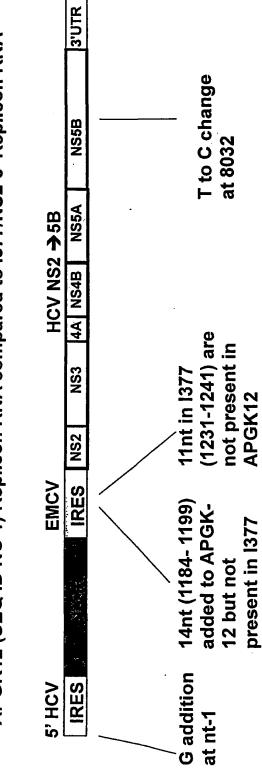
wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

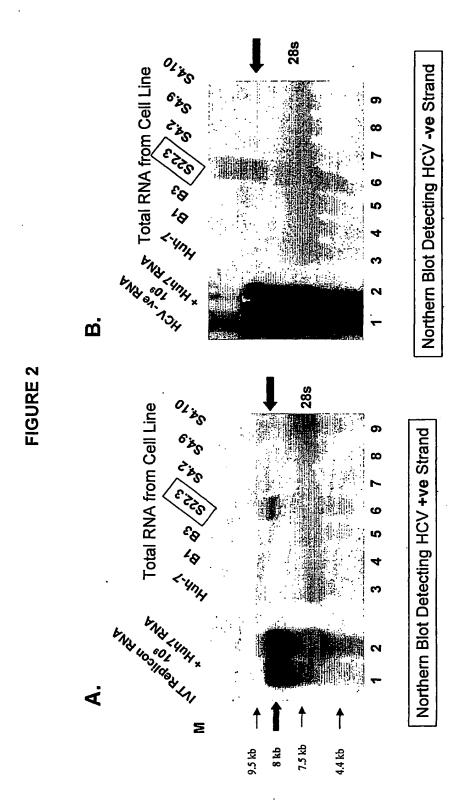
42. The method according to claim 41, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

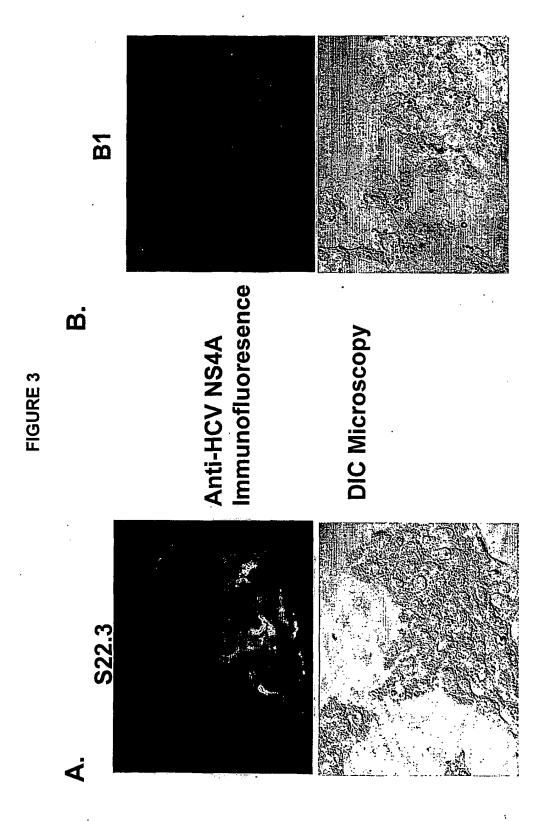
1/9

APGK12 (SEQ ID NO 1) Replicon RNA compared to I377/NS2-3' Replicon RNA

FIGURE 1







SUBSTITUTE SHEET (RULE 26)

3/9



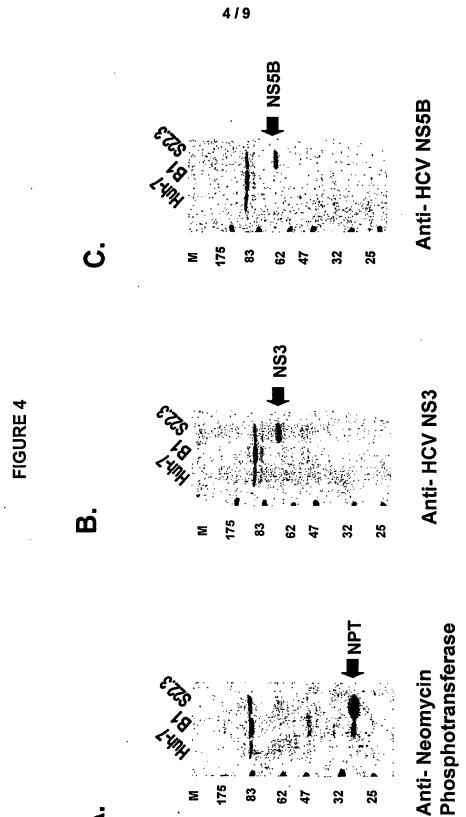


FIGURE 5A

	S 22-3 SEQ ID NO 2	R3 SEQ ID NO. 4	R3-rep SEQ ID NO. 7	R7 SEQ ID NO. 5	R16 SEQ ID NO 6
5'end - FIRST nt (HCV IRES)	*G (nt 1) A	G (nt 1) A	•	•	G (nt 1) A
Neo	•	A (nt 481) G	•		
EMCV IRES	•	A (nt 1739) G	•	•	•
NS.2		•			•
NS 3	•	G (nt 2778) A A (nt 2840) C A (nt 4062) G	T (nt 2509) C G (nt 2778) A A (nt 2840) C T (nt 3574) C A (nt 4052) G	A (nt 2935) G A (nt 2979) G	A (nt 2816) G A (nt 2979) G
NS 4A	A (nt 4446) R	A (nt 4448) G	C (nt 4387) T A (nt 4446) G C (nt 4507) T	•	C (nt 4475) T
NS 4B		T (nt 4855) C	T (nt 4865) C		
NS 5A	G (nt 6488) T A (nt 6268) R	A (nt 5351) G G (nt 5498) T	A (nt 5351) G G (nt 5498) T	A (nt 5324) G G (nt 5488) T	G (nt 5488) C T (nt 6320) C
		G (nt 5659) A C (nt 5871) T A (nt 6268) G	G (nt 5659) A T (nt 5838) C C (nt 5871) T A (nt 6115) G	.T (nt 6001) C	T (nt 6584) C
NS 5B	٠	A (nt 6662) G	•	C (nt 7252) T T (nt 8349) C	•
3'end - last 98 nt		•	:	•	•

first nt = G from HCV fres

6/9

G (2042) R S (2404) P S (1148) G E (1202) G R16 SEQ ID NO. 6 L (1701) F G (nt 1) A R3 Rep SEQ ID R7 SEQ ID NO. 5 E (1202) G G (2042) C M (2992) T I (1984) V T (1993) A G (2042) C L (2155) P P (2166) L R (1135) K S (1560) G K (1691) R R3 SEQ ID NO. 4 R (1135) K S (1560) G T (1993) A G (2042) C FIGURE 5B K (1691) R G (nt 1) A P (2166) L S 22-3 SEQ ID K (1691) mix K/R G (2042) C G (nt 1) A NO. 2 5'end - FIRST nt (HCV IRES) 3'end - last 98 nt NS 4A NS 4B NS 5A NS 5B NS₂ NS 3

first a.a. of NS2 = 810

7/9

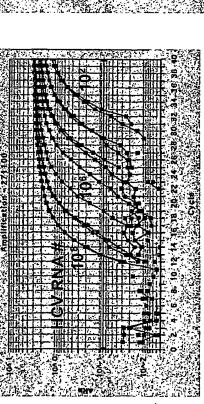
3'HCV UTR NS5B G(2042)C L(2155)P P(2166)L L(2155)P T(1993)A G(2042)C T(1993)A P(2166)L NS5A **AMINO ACID SUBSTITUTIONS** NS4B HCV NS2→5B K(1691)R K(1691)R R(1135)K S(1560)G FIGURE 6 R(1135)K S(1560)G NS3 NS2 EMCV IRES NeoR SEQ ID NO 24 SEQ ID NO 25 SEQ ID NO 7 SEQ ID NO 1 APGK-12 5' HCV IRES CLONE A (nt1) G (nt1) R3 rep G(nt1) A(nt1) 1100000cfu/µg 2000000cfu/µg 86 cfu/µg 77 cfu/µg

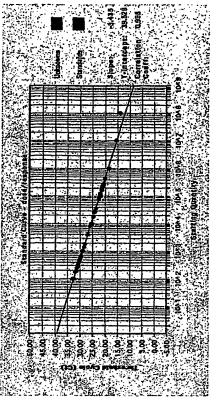
SUBSTITUTE SHEET (RULE 26)

8/9

HCV-Replicon: RNA Quantification

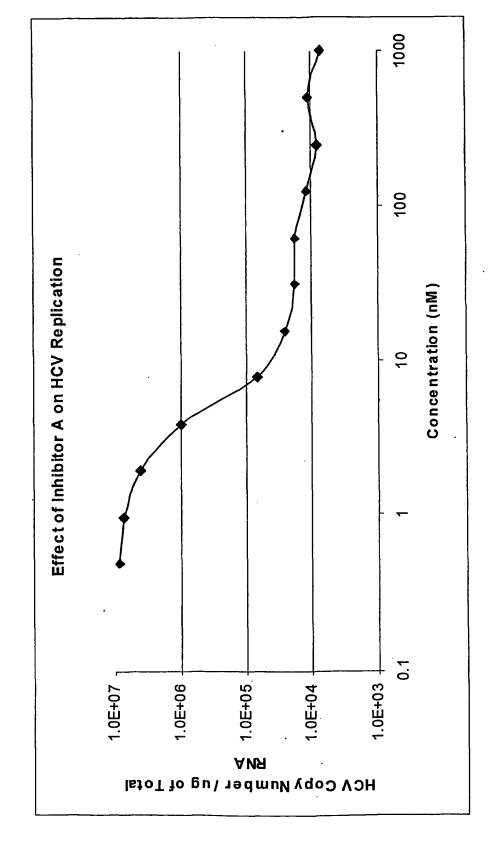
FIGURE 7





Starting RNA Quantity Ct = Threshold cycle α

FIGURE 8



1 / 93

SEQUENCE LISTING

```
<110> BOEHRINGER INGELHEIM (CANADA) LTD.
<120> SELF REPLICATING RNA MOLECULE FROM
  HEPATITIS C VIRUS
<130> 13/083
<150> 60/257,857
<151> 2000-12-22
<160> 25
<170> FastSEQ for Windows Version 4.0
<210> 1
<211> 8639
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1803)...(8408)
<400> 1
ggccagcccc cgattggggg cgacactcca ccatagatca ctcccctgtg aggaactact 60
gtcttcacgc agaaagcgtc tagccatggc gttagtatga gtgtcgtgca gcctccagga 120
cocccctcc cgggagagcc ataqtqqtct qcqqaaccqq tqaqtacacc qqaattqcca 180
ggacgacegg gteetttett ggatcaacec getcaatgee tggagatttg ggegtgeece 240
egegagactg ctageegagt agtgttgggt egegaaagge ettgtggtac tgeetgatag 300
ggtgcttgcg agtgccccgg gaggtctcgt agaccgtgca ccatgagcac gaatcctaaa 360
cctcaaagaa aaaccaaagg gcgcgccatg attgaacaag atggattgca cgcaggttct 420
ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac aatcggctgc 480
tctgatgccg ccgtgttccg gctgtcagcg caggggcgcc cggttctttt tgtcaagacc 540
gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc 600
acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggactgg 660
ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag 720
aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc ggctacctgc 780
ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat ggaagccggt 840
cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc 900
gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca tggcgatgcc 960
tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg 1020
ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat tgctgaagag 1080
cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc tcccgattcg 1140
cagcgcatcg ccttctatcg ccttcttgac gagttcttct gagttcgcgc ccagatgtta 1200
acagaccaca acggtttccc tctagcggga tcaattccgc cccccccct aacgttactg 1260
gccgaagccg cttggaataa ggccggtgtg cgtttgtcta tatgttattt tccaccatat 1320
tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg acgagcattc 1380
ctaggggtet tteceetete gecaaaggaa tgeaaggtet qttqaatqte qtgaaqgaaq 1440
cagttcctct ggaagettet tgaagacaaa caacgtetgt agegaceett tgeaggeage 1500
ggaaccccc acctggcgac aggtgcctct gcggccaaaa qccacqtqta taaqatacac 1560
ctgcaaaggc ggcacaaccc cagtgccacg ttgtgagttg gatagttgtg gaaagagtca 1620
aatggctctc ctcaagcgta ttcaacaagg ggctgaagga tgcccagaag gtaccccatt 1680
gtatgggatc tgatctgggg cctcggtgca catgctttac atgtgtttag tcgaggttaa 1740
aaaacgtcta ggccccccga accacgggga cgtggttttc ctttgaaaaa cacgataata 1800
```

cc atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val 1 5 10	1847
ggt ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct Gly Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala 20 25 30	1895
agg ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac Arg Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His 35 40 45	1943
ttg caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc Leu Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala 50 55 60	1991
gtc atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc Val Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile 65 70 75	2039
acc aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct Thr Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala 80 85 90 95	2087
ggt ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg 100 105 110	2135
gca tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg Ala Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met 115 120 125	2183
gct ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat Ala Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His 130 135 140	2231
ctc acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg Leu Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala 145 150 155	2279
gtg gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc Val Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile 160 165 170 175	2327
acc tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu 180 185 190	2375
ccc gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac Pro Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp 195 200 205	2423
agc ctt gaa ggg cag ggg tgg cga ctc ctc gcg cct att acg gcc tac Ser Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr 210 215 220	2471

	caa Gln 225															2519
	cgg Arg															2567
	aca Thr															2615
	tat Tyr															2663
atc Ile	acc Thr	caa Gln 290	atg Met	tac Tyr	acc Thr	aat Asn	gtg Val 295	gac Asp	cag Gln	gac Asp	ctc Leu	gtc Val 300	Gly	tgg Trp	caa Gln	2711
	ccc Pro 305															2759
	ctt Leu															2807
	Gly															2855
	aag Lys															2903
	gly ggc															2951
Val	gac Asp 385	Phe		Pro	Val		Ser			Thr		Met				2999
gtc Val 400	ttc Phe	acg Thr	gac Asp	aac Asn	tcg Ser 405	tcc Ser	cct Pro	ccg Pro	gcc Ala	gta Val 410	ccg Pro	cag Gln	aca Thr	ttc Phe	cag Gln 415	3047
	gcc Ala															3095
ccg Pro	gct Ala	gcg Ala	tat Tyr 435	gca Ala	gcc Ala	caa Gln	gly ggg	tat Tyr 440	aag Lys	gtg Val	ctt Leu	gtc Val	ctg Leu 445	aac Asn	ccg Pro	3143

		acc Thr								319 1
		aac Asn								3239
		tac Tyr								3287
		gcc Ala 500								3335
		act Thr								3383
 _	_	 gcg Ala	_	_		_	_	_		3431
		gtg Val								3479
		atc Ile								3527
_		 agg Arg 580			_		_	_	_	3575
		gcg Ala								3623
		ctt Leu								3671
		acg Thr								3719
		gac Asp								3767
		acc Thr 660								3815

											ggt Gly					3863
											ccc Pro					3911
											ggc Gly 715					3959
											cgg Arg					4007
											gag Glu					4055
											ttc Phe					4103
											gca Ala					4151
											tgg Trp 795					4199
											ely aaa					4247
											act Thr					4295
		Lys		Ile	Met		Cys	Met	Ser		gac Asp	Leu				4343
						Val					gca Ala					4391
											ggc Gly 875					4439
tcc Ser 880	gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	4487

										ctc Leu						4535
										aag Lys						4583
										gct Ala						4631
										aag Lys						4679
										tcc Ser 970						4727
										gcc Ala						4775
				His			_		Asn	atc Ile	_			Trp		4823
	gcc	caa	ctt	gct	cct	ccc	agc	gct	gct	tct	act	ttc	qta	aac	acc	4871
АТа	Ala	Gln 1010	Leu		Pro	Pro		Ala		Ser			Val			
ggc	atc	1010 gct Ala	Leu) gga	Ala gcg	gct.	gtt	Ser 1019 ggc Gly	Ala	Ala ata		Ala	Phe 1020 ggg Gly	Val) aag	Gly gtg	Ala	4919
ggc Gly gtg	atc Ile 1025 gat Asp	gct Ala	Leu gga Gly	Ala gcg Ala gca	gct. Ala ggt	gtt Val 1030 tat Tyr	Ser 1019 ggc Gly)	Ala agc Ser	Ala ata Ile	ser ggc Gly	ctt Leu 1035 gca Ala	gly aga 1050	Val aag Lys gcg	Gly gtg Val	Ala ctt Leu gtg	
ggc Gly gtg Val 1040	atc Ile 1025 gat Asp	gct Ala att Ile	gga Gly ttg Leu gtc Val	gcg Ala gca Ala	gct. Ala ggt Gly 1045 agc Ser	gtt Val 1030 tat Tyr	ggc Gly gga Gly	Ala agc Ser gca Ala atg Met	Ala ata Ile ggg Gly	ggc Gly gtg Val 1050 tcc ser	ctt Leu 1035 gca Ala	999 Gly ggc Gly	val aag Lys gcg Ala gac Asp	gtg Val ctc Leu	ctt Leu gtg Val 1055 gtt Val	
ggc Gly gtg Val 1040 gcc Ala	atc Ile 1025 gat Asp ttt Phe	gct Ala att Ile aag Lys	gga Gly ttg Leu gtc Val	gca Ala gca Ala atg Met 1060 gct	gct. Ala ggt Gly 1045 agc Ser	gtt Val 1030 tat Tyr ggc Gly	ggc Gly gga Gly gag Glu tcc	agc Ser gca Ala atg Met	ata Ile ggg Gly ccc Pro 1069	ggc Gly gtg Val 1050 tcc ser	ctt Leu 1035 gca Ala) acc Thr	Phe 1020 ggg Gly gag Glu gag Glu	val aag Lys gcg Ala gac Asp	gtg Val ctc Leu ctg Leu 1070	ctt Leu gtg Val 1055 gtt Val	4967
ggc Gly gtg Val 1040 gcc Ala aac Asn	atc Ile 1025 gat Asp ttt Phe cta Leu	gct Ala att Ile aag Lys ctc Leu gca	gga Gly ttg Leu gtc Val cct Pro 1079	gca Ala gca Ala atg Met 1060 gct Ala	gct. Ala ggt Gly 1045 agc Ser atc Ile	gtt Val 1030 tat Tyr ggc Gly ctc Leu	ggc Gly gga Gly gag Glu tcc ser	agc Ser gca Ala atg Met cct Pro 1080	ata Ile ggg Gly ccc Pro 1069 ggc Gly	ggc Gly gtg Val 1050 tcc ser 6	ctt Leu 1035 gca Ala acc Thr cta Leu	Phe 1020 ggg Gly ggc Gly gag Glu gtc Val	aag Lys gcg Ala gac Asp gtc Val 1089	gtg Val ctc Leu ctg Leu 1070 ggg Gly	ctt Leu gtg Val 1055 gtt Val gtc Val	4967 5015

								7 /	93		•					
gtc Val 1120	tcc Ser	ccc Pro	acg Thr	cac His	tat Tyr 1129	Val	cct Pro	gag Glu	agc Ser	gac Asp 1130	Ala	gca Ala	gca Ala	cgt Arg	gtc Val 1135	5207
	cag Gln				Ser					Gln					Leu	5255
	cag Gln			Asn					Thr					Ser		5303
	aga Arg		Val					Cys					Asp			5351
	tgg Trp 1185	Leu					Leu					Gly				5399
	tca Ser					Tyr					Arg					5447
	caa Gln				Pro					Ile					Lys	5495
	ggt Gly			Arg					Arg					Thr		5543
	gga Gly		Phe					Tyr					Cys			5591
	ccg Pro 1265	Ala					Arg					Val				5639
	tac Tyr					Arg					His					5687
atg Met	acc Thr	act Thr	gac Asp	aac Asn 1300	Val	aag Lys	tgc Cys	ccg Pro	tgt Cys 130!	${\tt Gln}$	gtt Val	ccg Pro	gcc Ala	ccc Pro 1310	Glu	5735
ttc Phe	ttc Phe	aca Thr	gaa Glu 1315	Val	gat Asp	ej aaa	gtg Val	cgg Arg 1320	Leu	cac His	agg Arg	tac Tyr	gct Ala 132	Pro	gcg Ala	5783
tgc Cys	aaa Lys	ccc Pro 1330	Leu	cta Leu	cgg Arg	gag Glu	gag Glu 1335	Val	aca Thr	ttc Phe	ctg Leu	gtc Val 134	Gly	ctc Leu	aat Asn	5831

		Leu	_			Gln 1350	Leu		-			Glu	_	_	_	5679
_	Val				_	ctc Leu		_			His		_			5927
					Leu	gcc Ala				Pro					Ser	5975
				${\tt Gln}$		tct Ser			Ser					Cys		6023
	_		Asp		_	gac Asp	_	Asp				_	Asn		_	6071
		Gln				ggg Gly 1430	Asn					Glu				6119
	Val					tct Ser					Gln					6167
					Val	ccg Pro				Leu					Lys	6215
		_		Met		ata Ile		_	Arg	_	_			Pro		6263
			Ser			gac Asp		Asp					Val			6311
	_	Pro	_	_		gcc Ala 1510	Lys	_		_		Pro				6359
	Lys					ctg Leu 5					Val					6407
					Lys	acc Thr				Ser					Val	6455
				Ala		gcc Ala	Ser		Asp					Asp		6503

		Ser		gtt Val			Tyr					Pro			6551
	Pro			ccc Pro	_	Leu	_				Trp			_	6599
Glu				gag Glu 1605	qaA					Ser					6647
				atc Ile)					Ala					Leu	6695
			Leu	agc Ser			_	Leu	_				Leu	_	6743
		Thr		cgc Arg			Ser					Lys			6791
	Arg			gtc Val		Asp					Asp				6839
Met				gcg Ala 1685	Ser					Lys					6887
				ctg Leu)					Ser					Phe	6935
			Lys	gac Asp				Leu					Val	aac Asn	6983
		Ser		tgg Trp			Leu					Glu			·7031
	Thr			atg Met		Lys					Суѕ				7079
Lys				aag Lys 1765	Pro					Val					7127
				gag Glu)					Tyr					Thr	7175

.10 / 93

ctc cct cag gcc g Leu Pro Gln Ala N 1795			Gly Phe Gln T	
gga cag cgg gtc g Gly Gln Arg Val (1810	Glu Phe Leu V			
cct atg ggc ttc g Pro Met Gly Phe A 1825	-		_	
gag aat gac atc o Glu Asn Asp Ile A 1840		Slu Ser Ile	_	
gcc ccc gaa gcc a Ala Pro Glu Ala A			Leu Thr Glu A	
atc ggg ggc ccc of the Gly Gly Pro 1875	-		Gln Asn Cys G	-
cgg tgc cgc gcg a Arg Cys Arg Ala 8 1890	Ser Gly Val I			
aca tgt tac ttg a Thr Cys Tyr Leu I 1905				
gac tgc acg atg o Asp Cys Thr Met I 1920				
agc gcg ggg acc o Ser Ala Gly Thr o			Leu Arg Ala E	
gct atg act aga t Ala Met Thr Arg 1 1955	Tyr Ser Ala P	Pro Pro Gly	gac ccg ccc a Asp Pro Pro I	ys Pro Glu
tac gac ttg gag t Tyr Asp Leu Glu 1 1970	Leu Ile Thr S			
cac gat gca tct g His Asp Ala Ser (1985				
acc ccc ctt gcg of Thr Pro Leu Ala 2				

11/93

	ac atc atc atg tat gcg sn Ile Ile Met Tyr Ala 2025		₹5
	ct cat ttc ttc tcc atc hr His Phe Phe Ser Ile 2040		13
	ta gat tgt cag atc tac eu Asp Cys Gln Ile Tyr 2055)1
	ta cct cag atc att caa eu Pro Gln Ile Ile Gln 2070		39
Ser Ala Phe Ser Leu Hi	at agt tac tct cca ggt is Ser Tyr Ser Pro Gly 085 2090	Glu Ile Asn Arg Val	37
	aa ctt ggg gta ccg ccc ys Leu Gly Val Pro Pro 2105		35
	tc cgc gct agg cta ctg al Arg Ala Arg Leu Leu 2120		33
	ag tac ctc ttc aac tgg ys Tyr Leu Phe Asn Trp 2135		31
	tc ccg gct gcg tcc cag le Pro Ala Ala Ser Gln 2150		79
Trp Phe Val Ala Gly Ty	ac agc ggg gga gac ata yr Ser Gly Gly Asp Ile 165 2170	Tyr His Ser Leu Ser	27
	gg ttc atg tgg tgc cta rp Phe Met Trp Cys Leu 2185		75
ggg gta ggc atc tat ct Gly Val Gly Ile Tyr Le 2195	ta ctc ccc aac cga tga eu Leu Pro Asn Arg * 2200	acggggaget aaacacteca 842	28
ggccaatagg ccatcctgtt ttttttttt ttttctcctt	tttttccctt tttttttc ttttttcct cttttttcc gctagctgtg aaaggtccgt	ttttttttt tttttttt 848 ttttctttcc tttggtggct 854 gagccgcttg actgcagaga 860 863	48 08

<210> 2

<211> 8642

<212> DNA

<213> HCV

<220>

```
<221> CDS
<222> (1802)...(8407)
<221> variation
<222> 6268
\langle 223 \rangle r = a or q
<221> variation
<222> 4446
<223> r = a or g
<400> 2
accageceee gattggggge gacaeteeae catagateae teecetgtga ggaactaetg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gateaacccg etcaatgeet ggagatttgg gegtgeecce 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaaqaaa aaccaaaqqq cqcqccatqa ttqaacaaqa tqqattqcac qcaqqttctc 420
eggeegettg ggtggagagg ctattegget atgaetggge acaacagaca ateggetget 480
ctgatgeege egtgtteegg etgteagege aggggegeee ggttettttt gteaagaeeg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt teettgegca getgtgeteg acgttgteac tgaageggga agggactgge 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg cccgaeggeg aggatetegt egtgaeceat ggegatgeet 960
gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee cagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc ccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt teceeteteg ceaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatet gatetgggge eteggtgeac atgetttaca tgtgtttagt egaggttaaa 1740
aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac egg gag atg gea gea teg tge gga gge geg gtt tte gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc qct agg
                                                                   1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arq
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                   1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                   1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
```

												ttt Phe				2041
												ctc Leu				2089
												ctc Leu				2137
												gtc Val 125				2185
												tat Tyr				2233
												gac Asp				2281
gca Ala	gtt Val	gag Glu	ccc Pro	gtc Val 165	gtc Val	ttc Phe	tct Ser	gat Asp	atg Met 170	gag Glu	acc Thr	aag Lys	gtt Val	atc Ile 175	acc Thr	2329
tgg Trp	eja aaa	gca Ala	gac Asp 180	acc Thr	gcg Ala	gcg Ala	tgt Cys	999 Gly 185	gac Asp	atc Ile	atc Ile	ttg Leu	ggc Gly 190	ctg Leu	ccc Pro	2377
gtc Val	tcc Ser	gcc Ala 195	cgc Arg	agg Arg	eja aaa	agg Arg	gag Glu 200	ata Ile	cat His	ctg Leu	gga Gly	ccg Pro 205	gca Ala	gac Asp	agc Ser	2425
												acg Thr				2473
			Arg	Gly	Leu	Leu		Cys	Ile	Ile	Thr	agc Ser				2521
cgg Arg	gac Asp	agg Arg	aac Asn	cag Gln 245	gtc Val	gag Glu	ejy aaa	gag Glu	gtc Val 250	caa Gln	gtg Val	gtc Val	tcc Ser	acc Thr 255	gca Ala	2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	gcg Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	gtc Val	2617
tat Tyr	cat His	ggt Gly 275	gcc Ala	ggc Gly	tca Ser	aag Lys	acc Thr 280	ctt Leu	gcc Ala	ggc	cca Pro	aag Lys 285	gly	cca Pro	atc Ile	2665

				acc Thr												2713
				cgt Arg												2761
				acg Thr 325												2809
	_	_		ej aaa	_							_			_	2857
				gly												2905
				gct Ala												2953
				gtc Val												3001
				tcg Ser 405												3049
				gcc Ala												3097
				gcc Ala												3145
Val			Thr	cta Leu	Gly	Phe	${\tt Gly}$		Tyr	Met		Lys				3193
atc Ile 465	gac Asp	cct Pro	aac Asn	atc Ile	aga Arg 470	acc Thr	gjå aaa	gta Val	agg Arg	acc Thr 475	atc Ile	acc Thr	acg Thr	ggt Gly	gcc Ala 480	3241
				tcc Ser 485												3289
tct Ser	gjà aaa	ggc Gly	gcc Ala 500	tat Tyr	gac Asp	atc Ile	ata Ile	ata Ile 505	tgt Cys	gat Asp	gag Glu	tgc Cys	cac His 510	tca Ser	act Thr	3337

							atc Ile 520									3385
							gtg Val									3433
							aac Asn									3481
							ggc									3529
							ttc Phe									3577
							eoo Gly ggc									3625
							gtc Val									3673
							atg Met									3721
							tgt Cys									3769
							gag Glu	Thr								3817
~+~								665					670			
			Ser	Gln	Arg	Arg	ggc Gly 680	agg Arg	Thr	Gly	agg Arg	Gly	agg Arg			3865
Val att	Ser tac	Arg 675 agg	Ser ttt	Gln gtg	Arg	Arg cca	Gly	agg Arg gaa	Thr	ccc	agg Arg tcg	Gly 685 ggc	agg Arg atg	Met	Gly gat	3865 3913
val att Ile tcc	tac Tyr 690	Arg 675 agg Arg	Ser ttt Phe ctg	Gln gtg Val tgc	act Thr	cca Pro 695	Gly 680 gga	agg Arg gaa Glu	Thr cgg Arg	ggc ggc	agg Arg tcg Ser 700	Gly 685 ggc Gly	agg Arg atg Met	Met ttc Phe tac	Gly gat Asp gag	

	ttg ccc Leu Pro 740											4057
Phe Thr	ggc ctc Gly Leu 755		_	Ala			_		_		_	4105
	gga gac Gly Asp											4153
	agg gct Arg Ala											4201
_	ata cgg Ile Arg	_	_	_				_		_	_	4249
	ctg gga Leu Gly 820											4297
	tac atc Tyr Ile 835			Ser								4345
	tgg gtg Trp Val											4393
	aca aca Thr Thr											4441
	ceg gcc Pro Ala											4489
	atg gaa Met Glu 900	Glu Cys		His	Leu	Pro	Tyr	Ile		Gln		4537
	ctc gcc Leu Ala 915			Gln								4585
	acc aag Thr Lys											4633
	acc ctc Thr Leu											4681

	gjå aaa															4729
	ata Ile															4777
acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	G1Å BBB	gga Gly 100	Trp	gtg Val	gcc Ala	4825
	caa Gln 1010	Leu					Ala					Val				4873
atc Ile 1025	gct Ala	gga Gly	gcg Ala	gct Ala	gtt Val 1030	Gly	agc Ser	ata Ile	Gly	ctt Leu 1035	Gly	aag Lys	gtg Val	ctt Leu	gtg Val 1040	4921
	att Ile				Tyr					Ala					Ala	4969
	aag Lys			Ser					Ser					Val		5017
cta Leu	ctc Leu	cct Pro 1075	Ala	atc Ile	ctc Leu	tcc Ser	cct Pro 1080	Gly	gcc Ala	cta Leu	gtc Val	gtc Val 1085	Gly	gtc Val	gtg Val	5065
tgc Cys	gca Ala 1090	Ala	ata Ile	ctg Leu	cgt Arg	cgg Arg 1095	His	gtg Val	ggc Gly	cca Pro	999 Gly 1100	Glu	gly aaa	gct Ala	gtg Val	5113
cag Gln 1105	tgg Trp	atg Met	aac Asn	cgg Arg	ctg Leu 1110	Ile	gcg Ala	ttc Phe	gct Ala	tcg Ser 1119	Arg	ggt Gly	aac Asn	cac His	gtc Val 1120	5161
tcc Ser	ccc Pro	acg Thr	cac His	tat Tyr 1125	Val	cct Pro	gag Glu	agc Ser	gac Asp 1130	Ala	gca Ala	gca Ala	cgt Arg	gtc Val 1135	Thr	5209
cag Gln	atc Ile	ctc Leu	tct Ser 1140	Ser	ctt Leu	acc Thr	atc Ile	act Thr 1145	Gln	ctg Leu	ctg Leu	aag Lys	agg Arg 1150	Leu	cac His	5257
cag Gln	tgg Trp	atc Ile 1155	Asn	gag Glu	gac Asp	tgc Cys	tcc Ser 1160	Thr	cca Pro	tgc Cys	tcc Ser	ggc Gly 1169	Ser	tgg Trp	cta Leu	5305
aga Arg	gat Asp 1170	Val	tgg Trp	gat Asp	tgg Trp	Ile	tgc Cys	Thr	gtg Val	Leu	act Thr	Asp	ttc Phe	aag Lys	acc Thr	5353

tgg ctc cag to Trp Leu Gln Se 1185					5401
tca tgt caa cg Ser Cys Gln Ar			Arg Gly Asp		5449
caa acc acc to Gln Thr Thr Cy 12					5497
tgt tcc atg ag Cys Ser Met Ar 1235				Thr Trp His	5545
gga aca ttc co Gly Thr Phe Pr 1250		Tyr Thr Thr			5593
ccg gcg cca aa Pro Ala Pro As 1265					5641
tac gtg gag gt Tyr Val Glu Va			His Tyr Val		5689
acc act gac as Thr Thr Asp As 13					5737
ttc aca gaa gt Phe Thr Glu Va 1315				Pro Ala Cys	5785
aaa ccc ctc ct Lys Pro Leu Le 1330		Val Thr Phe			5833
tac ctg gtt gg Tyr Leu Val Gl 1345	y Ser Gln Leu	Pro Cys Glu	Pro Glu Pro		5881
gtg ctc act to Val Leu Thr Se			His Ile Thr		5929
gct aag cgt ag Ala Lys Arg Ar 13					5977
tca gct agc ca					6025

cgt cat gac tcc Arg His Asp Ser 1410		Asp Leu Il			6073
cgg cag gag atg Arg Gln Glu Met 1425					6121
gta gta att ttg Val Val Ile Leu		Glu Pro Le			6169
agg gaa gta tcc Arg Glu Val Ser 146	Val Pro Ala	gag atc ct Glu Ile Le 1465	g cgg agg tcc u Arg Arg Ser	agg aaa ttc Arg Lys Phe 1470	6217 [.]
cct cga gcg atg Pro Arg Ala Met 1475	ccc ata tgg Pro Ile Trp	gca cgc cc Ala Arg Pr 1480	g gat tac aac o Asp Tyr Asn 148	Pro Pro Leu	6265
ttr gag tcc tgg Xaa Glu Ser Trp 1490		Asp Tyr Va			6313
tgt cca ttg ccg Cys Pro Leu Pro 1505	cct gcc aag Pro Ala Lys 1510	gcc cct cc Ala Pro Pr	g ata cca cct o Ile Pro Pro 1515	cca cgg agg Pro Arg Arg 1520	6361
aag agg acg gtt Lys Arg Thr Val	gtc ctg tca Val Leu Ser 1525	Glu Ser Th	c gtg tct tct r Val Ser Ser 30	gcc ttg gcg Ala Leu Ala 1535	6409
gag ctc gcc aca Glu Leu Ala Thr 154	Lys Thr Phe				6457
agc ggc acg gca Ser Gly Thr Ala 1555				Asp Gly Asp	6505
gcg gga tcc gac Ala Gly Ser Asp 1570		Tyr Ser Se			6553
gag ccg ggg gat Glu Pro Gly Asp 1585	ccc gat ctc Pro Asp Leu 1590	agc gac gg Ser Asp Gl	g tot tgg tot y Ser Trp Ser 1595	acc gta agc Thr Val Ser 1600	6601
gag gag gct agt Glu Glu Ala Ser	gag gac gtc Glu Asp Val 1605	Val Cys Cy	c tog atg toc s Ser Met Ser	tac aca tgg Tyr Thr Trp 1615	6649
aca ggc gcc ctg Thr Gly Ala Leu . 162	atc acg cca Ile Thr Pro	tgc gct gc	g gag gaa acc	aag ctg ccc	6697
atc aat gca ctg Ile Asn Ala Leu 1635	agc aac tct Ser Asn Ser	ttg ctc cg Leu Leu Ar 1640	t cac cac aac g His His Asn 164!	Leu Val Tyr	6745

gct aca aca tct cgc agc gca agc ctg cgg cag aag aag gtc acc ttt Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe 1650 1655 1660	6793
gac aga ctg cag gtc ctg gac gac cac tac cgg gac gtg ctc aag gag Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu 1665 1670 1675 1680	6841
atg aag gcg aag gcg tcc aca gtt aag gct aaa ctt cta tcc gtg gag Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu 1685 1690 1695	6889
gaa gcc tgt aag ctg acg ccc cca cat tcg gcc aga tct aaa ttt ggc Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly 1700 1705 1710	6937
tat ggg gca aag gac gtc cgg aac cta tcc agc aag gcc gtt aac cac Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His 1715 1720 1725	6985
atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile 1730 1735 1740	7033
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu 1745 1750 1755 1760	7081
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly 1765 1770 1775	7129
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu 1780 1785 1790	7177
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly 1795 1800 1805	7225
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro 1810 1815 1820	7273
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu 1825 1830 1835 1840	7321
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala 1845 1850 1855	7369
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile 1860 1865 1870	7417

1875		t aaa ggg cag r Lys Gly Glr 1880		Tyr Arg Arg	7465
tgc cgc gcg ag Cys Arg Ala Se 1890	r Gly Val Le				7513
tgt tac ttg as Cys Tyr Leu Ly 1905	g gcc gct gc s Ala Ala Al 1910	g gcc tgt cga a Ala Cys Arg	a gct gcg aag g Ala Ala Lys 1915	ctc cag gac Leu Gln Asp 1920	7561
tgc acg atg ct Cys Thr Met Le			val Val Ile		7609
geg ggg acc ca Ala Gly Thr Gl					7657
atg act aga ta Met Thr Arg Ty 1955	c tct gcc cc r Ser Ala Pr	c cct ggg gad o Pro Gly Asp 1960	c ccg ccc aaa o Pro Pro Lys 1969	Pro Glu Tyr	7705
gac ttg gag tt Asp Leu Glu Le 1970	g ata aca tc u Ile Thr Se 19	r Cys Ser Ser	e aat gtg tca Asn Val Ser 1980	gtc gcg cac Val Ala His	7753
gat gca tct gg		g tac tat cto	acc cgt gac	ccc acc acc	7801
Asp Ala Ser Gl	y Lys Arg Va 1990	l Tyr Tyr Leu			7001
	1990 g gct gcg tg	g gag aca gct	Thr Arg Asp 1995 aga cac act Arg His Thr	Pro Thr Thr 2000	7849
ccc ctt gcg cg Pro Leu Ala An tcc tgg cta gg Ser Trp Leu Gl	1990 g gct gcg tg g Ala Ala Tr 2005 c aac atc at	g gag aca gct p Glu Thr Ala 201 c atg tat gcg	Thr Arg Asp 1995 aga cac act A Arg His Thr 10	Pro Thr Thr 2000 cca gtc aat Pro Val Asn 2015 tgg ġca agg	
ccc ctt gcg cg Pro Leu Ala An tcc tgg cta gg Ser Trp Leu Gl	g gct gcg tgg Ala Ala Tr 2005 c aac atc at y Asn Ile Il 20 g act cat tt	g gag aca gct p Glu Thr Ala 201 c atg tat gcc e Met Tyr Ala 2025 c ttc tcc atc	Thr Arg Asp 1995 aga cac act Arg His Thr 10 g ccc acc ttg A Pro Thr Leu c ctt cta gct	Pro Thr Thr 2000 cca gtc aat Pro Val Asn 2015 tgg ġca agg Trp Ala Arg 2030 cag gaa caa Gln Glu Gln	7849
ccc ctt gcg cc Pro Leu Ala Ar tcc tgg cta gg Ser Trp Leu Gl 20 atg atc ctg at Met Ile Leu Me	g gct gcg tgg Ala Ala Tr 2005 c aac atc at y Asn Ile Il 20 g act cat tt t Thr His Ph	g gag aca gct p Glu Thr Ala 201 c atg tat gcg e Met Tyr Ala 2025 c ttc tcc atc e Phe Ser Ile 2040 t cag atc tac s Gln Ile Tyr	Thr Arg Asp 1995 aga cac act Arg His Thr 10 g ccc acc ttg A Pro Thr Leu c ctt cta gct Leu Leu Ala 204!	Pro Thr Thr 2000 cca gtc aat Pro Val Asn 2015 tgg ġca agg Trp Ala Arg 2030 cag gaa caa Gln Glu Gln tac tcc att	7849 7897
ccc ctt gcg cc Pro Leu Ala Ar tcc tgg cta gg Ser Trp Leu Gl atg atc ctg at Met Ile Leu Me 2035 ctt gaa aaa gc Leu Glu Lys Al	g gct gcg tgg Ala Ala Tr 2005 c aac atc at y Asn Ile Il 20 g act cat tt t Thr His Ph c cta gat tg a Leu Asp Cy 20 c cta cct ca	g gag aca gct p Glu Thr Ala 201 c atg tat gcg e Met Tyr Ala 2025 c ttc tcc atc e Phe Ser Ile 2040 t cag atc tac s Gln Ile Tyr 55	Thr Arg Asp 1995 aga cac act Arg His Thr 10 g ccc acc ttg Pro Thr Leu c ctt cta gct Leu Leu Ala 2049 c ggg gcc tgt Gly Ala Cys 2060 a cga ctc cac	Pro Thr Thr 2000 cca gtc aat Pro Val Asn 2015 tgg gca agg Trp Ala Arg 2030 cag gaa caa Gln Glu Gln tac tcc att Tyr Ser Ile	7849 7897

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	8137
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	8185
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	8233
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	8281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	8329
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190	8377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	8427
ggccaatagg ccatcetgtt tttttcccct ttttttttt tttttttt ttttttt	8547
<210> 3 <211> 2201 <212> PRT <213> HCV	
CZIS NCV	
<220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg	
<220> <221> VARIANT <222> 882	
<pre><220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg <221> VARIANT <222> 1489 <223> Xaa is Leu <400> 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly</pre>	
<pre><220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg <221> VARIANT <222> 1489 <223> Xaa is Leu <400> 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	
<pre><220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg <221> VARIANT <222> 1489 <223> Xaa is Leu <400> 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	
<pre><220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg <221> VARIANT <222> 1489 <223> Xaa is Leu <400> 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	

_		_	_			_		_	_			_			
-				85			_		90				Gln	95	•
Ile	Thr	Lys	Val	Pro	Tyr	Phe	Val	Arg 105	Ala	His	Gly	Leu	Ile 110	Arg	Ala
Сув	Met	Leu 115	Val	Arg	Lys	Val	Ala 120	GJA	GJA	His	Tyr	Val 125	Gln	Met	Ala
Leu	Met 130		Leu	Ala	Ala	Leu 135		Gly	Thr	Tyr	Val		Asp	His	Leu
		Leu	Arg	Asp	_		His	Ala	Gly			qaA	Leu	Ala	
145 Ala	Val	Glu	Pro		150 Val	Phe	Ser	Asp		155 Glu	Thr	Lys	Val		160 Thr
Trp	Gly	Ala	-	165 Thr	Ala	Ala	Cys	_	170 Asp	Ile	Ile	Leu	Gly	175 Leu	Pro
Val	Ser		180 Arg	Arg	Gly	Arg		185 Ile	His	Leu	Gly		190 Ala	Asp	Ser
Leu	Glu	195 Gly	Gln	Gly	Trp	Arg	200 Leu	Leu	Ala	Pro	Ile	205 Thr	Ala	Tyr	Ser
	210					215					220				_
225			_	_	230		_			235			Leu		240
Arg	Asp	Arg	Asn	Gln 245	Val	Glu	Gly	Glu	Val 250	Gln	Val	Val	Ser	Thr 255	Ala
Thr	Gln	Ser	Phe 260	Leu	Ala	Thr	Суѕ	Val 265	Asn	Gly	Val	Cys	Trp 270	Thr	Val
Tyr	His	Gly 275	Ala	Gly	Ser	Lys	Thr 280	Leu	Ala	Gly	Pro	Lys 285	Gly	Pro	Ile
Thr	Gln 290	Met	Tyr	Thr	Asn	Val 295	Asp	Gln	Asp	Leu	Val 300	Gly	Trp	Gln	Ala
Pro 305	Pro	Gly	Ala	Arg	Ser 310	Leu	Thr	Pro	Суз	Thr 315	Cys	Gly	Ser	Ser	Asp 320
Leu	Tyr	Leu	Val	Thr 325	Arg	His	Ala	Asp	Val 330	Ile	Pro	Val	Arg	Arg 335	Arg
Gly	Asp	Ser	Arg 340		Ser	Leu	Leu	Ser 345		Arg	Pro	Val	Ser 350		Leu
Lys	Gly	Ser 355		GĻY	Gly	Pro	Leu 360		Сув	Pro	Ser	Gly 365	His	Ala	Val
Gly	Ile 370		Arg	Ala	Ala	Val 375		Thr	Arg	Gly	Val 380		Lys	Ala	Val
Asp 385		Val	Pro	Val	Glu 390		Met	Glu	Thr	Thr 395		Arg	Ser	Pro	Val
	Thr	Asp	Asn			Pro	Pro	Ala			Gln	Thr	Phe		
Ala	His	Leu		405 Ala	Pro	Thr	Gly		410 Gly	Lys	Ser	Thr	Lys	415 Val	Pro
Ala	Ala		420 Ala	Ala	Gln	Gly		425 Lys	Val	Leu	Val		430 Asn	Pro	Ser
Val		435 Ala	Thr	Leu	Gly	Phe	440 Gly	Ala	Tyr	Met	Ser	445 Lys	Ala	His	Gly
T 1.	450	D	7	~ 1 _	7	455	~1	77-3	7	mb	460	m\	m\	63.	. .
465	Asp	Pro	ASII	тте	470	THE	GTA	Val	Arg	475	TTE	Thr	Thr	GIY	480
	Ile	Thr	Tyr	Ser 485		Tyr	Gly	Lys	Phe 490		Ala	Asp	Gly	_	
Ser	Gly	Gly			Asp	Ile	Ile	Ile 505		Asp	Glu	Cys	His	495 Ser	Thr
Asp	Ser		500 Thr	Ile	Leu	Gly			Thr	Val	Leu	_	510 Gln	Ala	Glu
Thr	Ala 530	515 Gly	Ala	Arg	Leu		520 Val	Leu	Ala	Thr		525 Thr	Pro	Pro	Gly
	230					535					540				

24 / 93

Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser 550 555 Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile 570 Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Cys Asp 585 Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr 600 Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile 615 Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp 630 635 Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser 650 Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Val Pro Gln Asp Ala 665 Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly 680 685 Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp 695 700 Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu 710 715 Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr 730 725 Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val 745 Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys 760 Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val 775 Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys 790 795 Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu 810 Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr His Pro Ile 825 Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr 840 Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr 855 Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser 870 875 Gly Xaa Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe 890 Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly 905 Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln 920 Thr Ala Thr Lys Gln Ala Glu Ala Ala Pro Val Val Glu Ser Lys 935 940 Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile 950 955 Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro 965 970 Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser. Pro Leu 985 Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala 1000

1010		er Ala Ala 015	Ser Ala Pho	e Val Gly Ala Gly
			Gly Leu Gl	y Lys Val Leu Val 1040
Asp Ile Leu Ala	Gly Tyr Gl 1045	ly Ala Gly		y Ala Leu Val Ala 1055
Phe Lys Val Met		lu Met Pro 106		a Asp Leu Val Asn 1070
1075		1080		l Val Gly Val Val 1085
Cys Ala Ala Ile 1090		rg His Val 095	Gly Pro Gly	y Glu Gly Ala Val 00
Gln Trp Met Asn 1105	Arg Leu Il 1110	le Ala Phe	Ala Ser Arg	Gly Asn His Val
Ser Pro Thr His	Tyr Val Pr 1125	ro Glu Ser		Ala Arg Val Thr
Gln Ile Leu Ser		or Ile Thr 114		ı Lys Arg Leu His 1150
Gln Trp Ile Asn 1155	Glu Asp Cy			r Gly Ser Trp Leu 1165
Arg Asp Val Trp			Val Leu Th	r Asp Phe Lys Thr
Trp Leu Gln Ser				Val Pro Phe Phe
		ys Gly Val		y Asp Gly Ile Met 1215
Gln Thr Thr Cys	Pro Cys Gl	ly Ala Gln 122	Ile Thr Gl	y His Val Lys Asn 1230
	-			Asn Thr Trp His
Gly Thr Phe Pro			Thr Gly Pro	Cys Thr Pro Ser
1250		555	124	50
	Tyr Ser Ar	255 cg Ala Leu		l Ala Ala Glu Glu
Pro Ala Pro Asn 1265	Tyr Ser Ar 1270 Thr Arg Va	g Ala Leu	Trp Arg Val 1275 Phe His Ty	l Ala Ala Glu Glu 1280 r Val Thr Gly Met
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy	eg Ala Leu al Gly Asp ys Pro Cys	Trp Arg Val 1275 Phe His Ty 1290 Gln Val Pro	l Ala Ala Glu Glu 1280 c Val Thr Gly Met 1295 o Ala Pro Glu Phe
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy 0	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu	Trp Arg Val 1275 Phe His Ty: 1290 Gln Val Pro	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 lu Val Thr	Trp Arg Val 1275 Phe His Ty 1290 Gln Val Pro 5 His Arg Ty	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy 0 Asp Gly Va Arg Glu Gl	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 lu Val Thr	Trp Arg Vai 1275 Phe His Ty: 1290 Gln Val Pro 5 His Arg Ty: Phe Leu Vai	l Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 I Gly Leu Asn Gln
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy 0 Asp Gly Va Arg Glu Gl	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 lu Val Thr	Trp Arg Vai 1275 Phe His Ty: 1290 Gln Val Pro 5 His Arg Ty: Phe Leu Vai	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl 13 Ser Gln Le 1350	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 lu Val Thr 335 eu Pro Cys	Trp Arg Va. 1275 Phe His Ty: 1290 Gln Val Pro 5 His Arg Ty: Phe Leu Va. 134 Glu Pro Gli 1355	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 100 1 Pro Asp Val Ala
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy Asp Gly Va Arg Glu Gl 13 Ser Gln Le 1350 Met Leu Th 1365 Leu Ala Ar	g Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 tu Val Thr 335 eu Pro Cys	Trp Arg Va. 1275 Phe His Ty. 1290 Gln Val Pro 5 His Arg Ty. Phe Leu Va. 134 Glu Pro Gl. 1355 Ser His Ild 1370 Pro Pro Ser	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln O Pro Asp Val Ala 1360 Thr Ala Glu Thr
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 130 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser Ala Lys Arg Arg	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl 13 Ser Gln Le 1350 Met Leu Th 1365 Leu Ala Ar	al Gly Asp ys Pro Cys 130 al Arg Leu 1320 tu Val Thr 335 eu Pro Cys ar Asp Pro	Trp Arg Value 1275 Phe His Tyres 1290 Gln Val Pros His Arg Tyres Phe Leu Value 134 Glu Pro Glue 1355 Ser His Ilonary Pro Pro Ser	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 40 Pro Asp Val Ala 1360 Thr Ala Glu Thr 1375 Leu Ala Ser Ser
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 1300 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser Ala Lys Arg Arg 138 Ser Ala Ser Gln 1395 Arg His Asp Ser	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl 13 Ser Gln Le 1350 Met Leu Th 1365 Leu Ala Ar O Leu Ser Al	rg Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 Lu Val Thr 335 eu Pro Cys nr Asp Pro rg Gly Ser 138 La Pro Ser 1400 La Asp Leu	Trp Arg Va. 1275 Phe His Ty. 1290 Gln Val Pro His Arg Ty. Phe Leu Va. 134 Glu Pro Gl. 1355 Ser His Ild 1370 Pro Pro Ser Leu Lys Ala	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 40 Pro Asp Val Ala 1360 Thr Ala Glu Thr 1375 Leu Ala Ser Ser 1390 Thr Cys Thr Thr 1405 Asn Leu Leu Trp
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 1300 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser Ala Lys Arg Arg 138 Ser Ala Ser Gln 1395 Arg His Asp Ser 1410 Arg Gln Glu Met	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl Ser Gln Le 1350 Met Leu Th 1365 Leu Ala Ar O Leu Ser Al Pro Asp Al	al Gly Asp ys Pro Cys 130 al Arg Leu 1320 au Val Thr 335 au Pro Cys ar Asp Pro rg Gly Ser 138 a Pro Ser 1400 la Asp Leu	Trp Arg Val 1275 Phe His Ty: 1290 Gln Val Pro His Arg Ty: Phe Leu Val Glu Pro Gli 1355 Ser His Ild 1370 Pro Pro Se: Leu Lys Ala Ile Glu Ala	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 40 Pro Asp Val Ala 1360 Thr Ala Glu Thr 1375 Leu Ala Ser Ser 1390 Thr Cys Thr Thr 1405 Asn Leu Leu Trp
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 1300 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser Ala Lys Arg Arg 138 Ser Ala Ser Gln 1395 Arg His Asp Ser 1410 Arg Gln Glu Met	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl 1350 Met Leu Th 1365 Leu Ala Ar O Leu Ser Al Pro Asp Al Gly Gly As 1430	rg Ala Leu al Gly Asp ys Pro Cys 130 al Arg Leu 1320 au Val Thr 335 au Pro Cys ar Asp Pro rg Gly Ser 138 a Pro Ser 1400 la Asp Leu 115 sn Ile Thr	Trp Arg Value 1275 Phe His Tyree 1290 Gln Val Prosection Shis Arg Tyree Phe Leu Value 1355 Ser His Ille 1370 Pro Pro Ser 155 Leu Lys Alamar Val Glu Arg Val Glu 1435	Ala Ala Glu Glu 1280 Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 40 Pro Asp Val Ala 1360 Thr Ala Glu Thr 1375 Leu Ala Ser Ser 1390 Thr Cys Thr Thr 1405 Asn Leu Leu Trp 20 1 Ser Glu Asn Lys 1440
Pro Ala Pro Asn 1265 Tyr Val Glu Val Thr Thr Asp Asn 1300 Phe Thr Glu Val 1315 Lys Pro Leu Leu 1330 Tyr Leu Val Gly 1345 Val Leu Thr Ser Ala Lys Arg Arg 138 Ser Ala Ser Gln 1395 Arg His Asp Ser 1410 Arg Gln Glu Met 1425 Val Val Ile Leu	Tyr Ser Ar 1270 Thr Arg Va 1285 Val Lys Cy O Asp Gly Va Arg Glu Gl 13 Ser Gln Le 1350 Met Leu Th 1365 Leu Ala Ar O Leu Ser Al Pro Asp Al Gly Gly As 1430 Asp Ser Ph 1445	rg Ala Leu al Gly Asp rs Pro Cys 130 al Arg Leu 1320 al Val Thr 335 au Pro Cys ar Asp Pro rg Gly Ser 138 a Pro Ser 1400 la Asp Leu 115 sn Ile Thr	Trp Arg Val 1275 Phe His Ty 1290 Gln Val Pro His Arg Ty Phe Leu Val 336 Glu Pro Gli 1355 Ser His Ilo 1370 Pro Pro Ser Leu Lys Ala 142 Arg Val Gli 1435 Leu Gln Ala 1450	Ala Ala Glu Glu 1280 C Val Thr Gly Met 1295 Ala Pro Glu Phe 1310 Ala Pro Ala Cys 1325 Gly Leu Asn Gln 10 1 Pro Asp Val Ala 1360 Thr Ala Glu Thr 1375 C Leu Ala Ser Ser 1390 A Thr Cys Thr Thr 1405 A Asn Leu Leu Trp 20 1 Ser Glu Asn Lys

26 / 93

Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu 1475 1480 Xaa Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly 1495 1500 Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg 1510 1515 1520 Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala 1525 1530 Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp 1545 Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp 1560 Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly 1570 1575 1580 Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser ¹⁵⁹⁰ 1595 Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp 1605 1610 Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro 1620 1625 1630 Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr 1635 1640 1645 Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe 1660. 1650 1655 Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu 1670 1675 Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu 1685 1690 Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly 1700 1705 1710 Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His 1715 1720 1725 Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile 1735 1730 1740 Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu 1750 i755 Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly 1765 1770 1775 Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu 1780 1785 1790 Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly 1800 1795 Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro 1810 1815 1820 Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu 1830 1835 Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala 1850 1845 Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile 1860 1865 Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg 1880 1885 Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr 1895 1900 Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp 1915 1910 Cys Thr Met Leu Val Cys Gly Asp Leu Val Val Ile Cys Glu Ser 1925 1930

27 / 93

Ala Gly Thr Gln Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala 1940 1945 Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr 1955 1960 1965 Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His 1975 1980 Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr 1990 1995 Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn 2005 2010 Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg 2020 2025 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045 Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2055 2060 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2070 2075 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2120 2125 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2135 2140 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 Val Gly Ile Tyr Leu Leu Pro Asn Arg 2195 2200 <210> 4 <211> 8643 <212> DNA <213> HCV <220> <221> CDS <222> (1802)...(8407) accagecece gattggggge gacactecae catagateae teceetgtga ggaactactg 60 tetteaegea gaaagegtet ageeatggeg ttagtatgag tgtegtgeag cetecaggae 120 cccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180 gacgaccggg teetttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecce 240 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420 cggccgcttg ggtggagagg ctattcggct atgactgggc gcaacagaca atcggctgct 480 ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600 cgacgggcgt teettgcgca getgtgctcg acgttgtcac tgaagcggga agggactggc 660 tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720 aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780

cattegacea ceaagegaaa categeateg agegageaeg tacteggatg gaageeggte 84 ttgtegatea ggatgatetg gacgaagage ateagggget egegeeagee gaactgtteeg 96 ceaggeteaa ggegegeatg eeegacgeg aggatetegt egtgaceat ggegatgeet 96 gettgeegaa tateatggtg gaaaatggee getttetegg atteategae tgtggeegge 16 tgggtgtgge ggacegetat eaggacatag egttggetae eegtgatatt getgaagage 16 ttggeggega atgggetgae egetteeteg tgetttaegg tategeegee eeegattege 11 agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 12 cagaceacaa eggttteeet etagegggat eaatteegee eeeceecta aegttaetgg 12 cegaageege ttggaataag geegggat gtttgtetat atgttattt eeaceatatt 13 geegtetttt ggeaatgtga gggeeeggaa acetggeeet gtettettga egageattee 13 taggggtett teeeeteteg eeaaaggaat geaaggtetg ttgaatgteg tgaaggaage 14 agtteetetg gaagettett gaagacaaae aacgtetgta gegaceettt geaggeageg 15 gaaceeceea eetggegaca ggtgeetetg eggeeaaaag eeaegtgtat aagataeaee 15 tgeaaaggeg geacaaceee agtgeeaegt tgtgagttgg atagttgg aaagagteaa 16 atggetetee teaagegtat teaacaaggg getgaaggat geeeagaagg taceecattg 16 atagggatet gatetgggge eteggteae atgetttaea tgtgtttagt egaggttaga 17 aaaacgtetag geeeceegaa eeaegggae gtggtttee tttgaaaaaae aegataatae 16 eatg gae egg gag atg gea gea teg tge gga gge geg gtt tte gta ggt 18 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1 5 10 15	00 00 00 00 00 00 00 00 00 00 00 00 00
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg 20 25 30	897
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg 19 Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu 35 40 45	945
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 19 Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val 50 55 60	993
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr 65 70 75 80	041
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly 85 90 95	089
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala 100 105 110	137
tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct 21 Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala 115 120 125	185
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu 130 135 140	233
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val 145 150 150 155 160	281

gca gtt gag Ala Val Glu					
tgg ggg gca Trp Gly Ala					
gtc tcc gcc Val Ser Ala 195					
ctt gaa ggg Leu Glu Gly 210					
caa cag acg Gln Gln Thr 225					
cgg gac agg Arg Asp Arg					
aca caa tct Thr Gln Ser	Phe Leu Ala 260	Thr Cys Val 265	Asn Gly Val	Cys Trp Thr 270	Val
tat cat ggt Tyr His Gly 275	Ala Gly Ser	Lys Thr Leu 280	Ala Gly Pro	Lys Gly Pro 285	Ile
acc caa atg Thr Gln Met 290	Tyr Thr Asn	Val Asp Gln 295	Asp Leu Val 300	Gly Trp Gln	Ala
ccc ccc ggg Pro Pro Gly 305	Ala Arg Ser 310	Leu Thr Pro	Cys Thr Cys 315	Gly Ser Ser	Asp 320
ctt tac ttg Leu Tyr Leu	gtc acg aag Val Thr Lys 325	cat gcc gat His Ala Asp	Val Ile Pro	gtg cgc cgg Val Arg Arg 335	cgg 2809 Arg
ggc gac agc Gly Asp Ser					
aag ggc tct Lys Gly Ser 355					
ggc atc ttt Gly Ile Phe 370	cgg gct gcc Arg Ala Ala	gtg tgc acc Val Cys Thr 375	cga ggg gtt Arg Gly Val 380	gcg aag gcg Ala Lys Ala	gtg 2953 Val

										act Thr 395						3001
										ccg Pro						3049
										aag Lys						3097
										ctt Leu						3145
										atg Met						3193
										acc Thr 475						3241
										ctt Leu						3289
										gat Asp						3337
										gtc Val						3385
										acc Thr						3433
										gag Glu 555						3481
										atc Ile						3529
aag Lys	Gly 999	gly aaa	agg Arg 580	cac His	ctc Leu	att Ile	ttc Phe	tgc Cys 585	cat His	tcc Ser	aag Lys	aag Lys	aaa Lys 590	tgt Cys	gat Asp	3577
gag Glu	ctc Leu	gcc Ala 595	gcg Ala	aag Lys	ctg Leu	tcc Ser	ggc Gly 600	ctc Leu	gga Gly	ctc Leu	aat Asn	gct Ala 605	gta Val	gca Ala	tat Tyr	3625

				_	gta Val		_				-		_	_		3673
					gct Ala 630											3721
					aat Asn											3769
				Phe	acc Thr											3817
gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	ggc Gly 680	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	Gly	3865
					act Thr											3913
					gag Glu 710											3961
					acc Thr											4009
					tgc Cys											4057
					cac His											4105
		Gly			ttc Phe		Tyr	Leu	Val	Ala		${\tt Gln}$				4153
tgc Cys 785					gct											4201
765	Ala	Arg	Ala	Gln	Ala 790	Pro	PIO	·	ser	795	Asp	0111	Mec		800	
tgt	ctc	ata	cgg	cta		cct	acg	ctg	cac	795 999	cca	acg	ccc	ctg	800 ctg	4249

acc aaa tac atc Thr Lys Tyr Ile 835			la Asp Leu G	
agc acc tgg gtg Ser Thr Trp Val 850				
tgc ctg aca aca Cys Leu Thr Thr 865				
gga agg ccg gcc Gly Arg Pro Ala		Asp Arg G	•	 Phe
gat gag atg gaa Asp Glu Met Glu 900	Glu Cys Ala			
atg cag ctc gcc Met Gln Leu Ala 915	_	_	ys Ala Ile (
aca gcc acc aag Thr Ala Thr Lys 930				 -
tgg cgg acc ctc Trp Arg Thr Leu 945			_	
agc ggg ata caa Ser Gly Ile Glm	_	Gly Leu S	_	 Pro
gcg ata gca tca Ala Ile Ala Ser 980	Leu Met Ala	_		
acc acc caa cat Thr Thr Gln His 995			le Leu Gly (
gcc caa ctt gct Ala Gln Leu Ala 1010		Ala Ala S		
atc gct gga gcg Ile Ala Gly Ala 1025				
gat att ttg gca Asp Ile Leu Ala		Ala Gly V		Ala

				acc gag gac Thr Glu Asp		
	Ala Ile 1		o Gly Ala	cta gtc gtc Leu Val Val 108	Gly Val	
tgc gca gcg Cys Ala Ala 1090	ata ctg o	cgt cgg ca Arg Arg Hi 1095	c gtg ggc s Val Gly	cca ggg gag Pro Gly Glu 1100	ggg gct Gly Ala	gtg 5113 Val
	Asn Arg I			tcg cgg ggt Ser Arg Gly 1115		
				gct gca gca Ala Ala Ala O		Thr
				ctg ctg aag Leu Leu Lys		
	Asn Glu A		r Thr Pro	tgc tcc ggc Cys Ser Gly 116	Ser Trp	
				ttg act gat Leu Thr Asp 1180		
	Ser Lys I			ccg gga gtc Pro Gly Val 1195		
				cgg ggc gac Arg Gly Asp O		Met
Gln Thr Thr	Cys Pro (a Gln Ile	acc gga cat Thr Gly His		
	Arg Ile V		Arg Thr	tgt agt aac Cys Ser Asn 124	Thr Trp	
		Asn Ala Ty		ggc ccc tgc Gly Pro Cys		
1230		1255		1260		

	gtg Val				Arg					His					Met	5689
	act Thr			Val					Gln					Glu		5737
ttc Phe	aca Thr	gaa Glu 131	Val	gat Asp	gj aaa	gtg Val	cgg Arg 1320	Leu	cac His	agg Arg	tac Tyr	gct Ala 1325	Pro	gcg Ala	tgc Cys	5785
	ccc Pro 1330	Leu					Val					Gly				5833
tac Tyr 1345	ctg Leu	gtt Val	gly aaa	tca Ser	cag Gln 1350	Leu	cca Pro	tgc Cys	gag Glu	ccc Pro 1355	Glu	ctg Leu	gac Asp	gta Val	gca Ala 1360	5881
gtg Val	ctc Leu	act Thr	tcc Ser	atg Met 1365	Leu	acc Thr	gac Asp	ccc Pro	tcc Ser 1370	His	att Ile	acg Thr	gcg Ala	gag Glu 1375	Thr	5929
	aag Lys			Leu					Pro					Ser		5977
tca Ser	gct Ala	agc Ser 1395	Gln	ctg Leu	tct Ser	gcg Ala	cct Pro 1400	Ser	ttg Leu	aag Lys	gca Ala	aca Thr 1405	Сув	act Thr	acc Thr	6025
cgt Arg	cat His 1410	Asp	tcc Ser	ccg Pro	gac Asp	gct Ala 1415	Asp	ctc Leu	atc Ile	gag Glu	gcc Ala 1420	Asn	ctc Leu	ctg Leu	tgg Trp	6073
cgg Arg 1425	cag Gln	gag Glu	atg Met	Gly ggc	999 Gly 1430	Asn	atc Ile	acc Thr	cgc Arg	gtg Val 1435	Glu	tca Ser	gaa Glu	aat Asn	aag Lys 1440	6121
gta Val	gta Val	att Ile	ttg Leu	gac Asp 1445	Ser	ttc Phe	gag Glu	ccg Pro	ctc Leu 1450	Gln	gcg Ala	gag Glu	gag Glu	gat Asp 1455	Glu	6169
agg Arg	gaa Glu	gta Val	tcc Ser 1460	Val	ccg Pro	gcg Ala	gag Glu	atc Ile 1465	Leu	cgg Arg	agg Arg	tcc Ser	agg Arg 1470	ГЛЗ	ttc Phe	6217
cct Pro	cga Arg	gcg Ala 1475	Met	ccc Pro	ata Ile	tgg Trp	gca Ala 1480	Arg	ccg Pro	gat Asp	tac Tyr	aac Asn 1485	Pro	cca Pro	ctg Leu	6265
	gag Glu 1490	Ser					Asp					Val				6313

	Pro	Leu				Lys					Pro					6361
aag Lys	agg Arg	acg Thr	gtt Val	gtc Val 152	Leu	tca Ser	gaa Glu	tct Ser	acc Thr 1530	Val	tct Ser	tct Ser	gcc Ala	ttg Leu 153!	Ala	6409
gag Glu	ctc Leu	gcc Ala	aca Thr 1540	Lys	acc Thr	ttc Phe	ggc Gly	agc Ser 154	Ser	gaa Glu	tcg Ser	tcg Ser	gcc Ala 155	Val	gac Asp	6457
agc Ser	ggć Gly	acg Thr 155!	Ala	acg Thr	gcc Ala	tct Ser	cct Pro 1560	Asp	cag Gln	ccc Pro	tcc Ser	gac Asp 156	Asp	ggc Gly	gac Asp	6505
gcg Ala	gga Gly 1570	tcc Ser	gac Asp	gtt Val	gag Glu	tcg Ser 157	Tyr	tcc Ser	tcc Ser	atg Met	ccc Pro 1580	Pro	ctt Leu	gag Glu	gl ^y aaa	6553
gag Glu 158!	Pro	gjå aaa	gat Asp	ccc Pro	gat Asp 1590	Leu	agc Ser	gac Asp	eja aaa	tct Ser 1599	Trp	tct Ser	aċc Thr	gta Val	agc Ser 1600	6601
gag Glu `	gag Glu	gct Ala	agt Ser	gag Glu 1609	Asp	gtc Val	gtc Val	tgc Cys	tgc Cys 1610	Ser	atg Met	tcc Ser	tac Tyr	aca Thr 1615	Trp	6649
acg Thr	gly	gcc Ala	ctg Leu 162(Ile	acg Thr	cca Pro	tgc Cys	gct Ala 1625	Ala	gag Glu	gaa Glu	acc Thr	aag Lys 1630	Leu	ccc Pro	6697
		gca Ala 1635	Leu					Leu					Leu			6745
gct Ala	aca Thr 1650	aca Thr	tct Ser	cgc Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	[.] 6793
gac Asp 1669	Arg	ctg Leu	cag Gln	gtc Val	ctg Leu 1670	Asp	gac Asp	cac His	tac Tyr	cgg Arg 1675	Asp	gtg Val	ctc Leu	aag Lys	gag Glu 1680	6841
atg Met	aag Lys	gcg Ala	aag Lys	gcg Ala 1685	Ser	aca Thr	gtt Val	aag Lys	gct Ala 1690	Lys	ctt Leu	cta Leu	tcc Ser	gtg Val 1695	Glu	6889
gaa Glu	gcc Ala	tgt Cys	aag Lys 1700	Leu	acg Thr	ccc Pro	cca Pro	cat His 1705	Ser	gcc Ala	aga Arg	tct Ser	aaa Lys 1710	Phe	ggc Gly	6937
tat Tyr	gly ggg	gca Ala 1715	Lys	gac Asp	gtc Val	Arg	aac Asn 1720	Leu	tcc Ser	agc Ser	aag Lys	gcc Ala 1725	Val	aac Asn	cac His	6985

		Ser				Asp 1735	Leu					Glu				7033
	Thr					aaa Lys)					Сув					7081
					Pro	gct Ala				Val					Gly	7129
				Glu		atg Met			Tyr					Thr		7177
			Val			tct Ser		Tyr					Ser			7225
		Val				gtg Val 1815	Asn					Lys				7273
	Gly					acc Thr					Ser					7321
					Glu	gag Glu				Gln					Ala	7369
	_	_	_	Gln	_	ata Ile		_	Leu					Tyr		7417
			Leu			tct Ser		Gly					Tyr			7465
_	_	Ala	_		-	ctg Leu 1895	Thr		_	-		Asn				7513
	Tyr					gcg Ala)					Ala					7561
					Cys	gga Gly				Val					Ser	7609
				Glu		gag Glu			Leu					Glu		7657

Met Thr Arg Tyr 1955			ecg ccc aaa cca Pro Pro Lys Pro 1965	
gac ttg gag ttg Asp Leu Glu Leu 1970	ata aca tca Ile Thr Ser 1975	Cys Ser Ser A	aat gtg tca gtc Asn Val Ser Val 1980	gcg cac 7753 Ala His
gat gca tct ggc Asp Ala Ser Gly 1985		Tyr Tyr Leu 1		
ccc ctt gcg cgg Pro Leu Ala Arg				
tcc tgg cta ggc Ser Trp Leu Gly 2020	Asn Ile Ile			Ala Arg
atg atc ctg atg Met Ile Leu Met 2035	Thr His Phe	Phe Ser Ile I 2040	Leu Leu Ala Gln 2045	Glu Gln
ctt gaa aaa gcc Leu Glu Lys Ala 2050		Gln Ile Tyr G		
gag cca ctt gac Glu Pro Leu Asp 2065		Ile Ile Gln A		
gca ttt tca ctc Ala Phe Ser Leu				
		2090	iu lle Asn Arg	Val Ala 2095
tca tgc ctc agg Ser Cys Leu Arg 2100	Lys Leu Gly	2090 gta ccg ccc t	tg cga gtc tgg	2095 aga cat 8137 Arg His
Ser Cys Leu Arg	Lys Leu Gly gtc cgc gct Val Arg Ala	gta ccg ccc t Val Pro Pro I 2105	tg cga gtc tgg Leu Arg Val Trp 2110	aga cat 8137 Arg His 0 agg gct 8185
Ser Cys Leu Arg 2100 cgg gcc aga agt Arg Ala Arg Ser	Lys Leu Gly gtc cgc gct Val Arg Ala aag tac ctc	gta ccg ccc t Val Pro Pro I 2105 agg cta ctg t Arg Leu Leu S 2120 ttc aac tgg g Phe Asn Trp A	Leu Arg Val Trp 2110 2110 2125 2125 2125	aga cat 8137 Arg His 0 agg gct 8185 Arg Ala aag ctc 8233
Ser Cys Leu Arg 2100 cgg gcc aga agt Arg Ala Arg Ser 2115 gcc act tgt ggc Ala Thr Cys Gly	Lys Leu Gly gtc cgc gct Val Arg Ala aag tac ctc Lys Tyr Leu 2135 atc ccg gct	gta ccg ccc t Val Pro Pro I 2105 agg cta ctg t Arg Leu Leu S 2120 ttc aac tgg g Phe Asn Trp A	ctg cga gtc tgg Leu Arg Val Trp 2110 ccc cag ggg ggg Ser Gln Gly Gly 2125 gca gta agg acc Ala Val Arg Thr 2140 ctg gat tta tcc	aga cat 8137 Arg His 0 agg gct 8185 Arg Ala aag ctc 8233 Lys Leu agc tgg 8281

38 / 93

8377 gee ega eee ege tgg tte atg tgg tge eta ete eta ett tet gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2185 2180 2190 8427 gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200 tttttttttt tttttttt ttttctttt tcccaatttt tttcctttc tttcctttgg 8547 tggctccatc ttagccctag tcacggctag ctgtgaaagg tccgtgagcc gcttgactgc 8607 agagagtgct gatactggcc tctctgcaga tcaagt <210> 5 <214> 8648 <212> DNA <213> HCV <220> <221> CDS <222> (1802)...(8407) <400> 5 gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60 tettcaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120 ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180 gacgaccggg teetttettg gatcaacccg etcaatgeet ggagatttgg gegtgeecee 240 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420 eggeegettg ggtggagagg etattegget atgaetggge acaacagaca ateggetget 480 etgatgeege egtgtteegg etgteagege aggggegeee ggttettttt gteaagaeeg 540 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600 cgacgggcgt teettgegea getgtgeteg acgttgteae tgaageggga agggaetgge 660 tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720 aagtateeat eatggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780 cattegacca ccaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840 ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900 ccaggeteaa ggegegeatg ceegaeggeg aggatetegt egtgaceeat ggegatgeet 960 gettgeegaa tateatggtg gaaaatggee gettttetgg atteategae tgtggeegge 1020 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080 ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140 agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200 cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260 ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320 gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380 taggggtett teceeteteg ecaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440 agtteetetg gaagettett gaagacaaac aaegtetgta gegaceettt geaggeageg 1500 gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620 atggetetee teaagegtat teaacaaggg getgaaggat geceagaagg taceceattg 1680 tatgggatet gatetgggge eteggtgeae atgetttaea tgtgtttagt egaggttaaa 1740 aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800 c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1.0

										ctc Leu 30			1897
	_							-		gca Ala		_	1945
										gat Asp			1993
									•	acc Thr			2041
										cag Gln			2089
		 _			_	_				att Ile 110		_	2137
										caa Gln			2185
										gac Asp			2233
										ctt Leu			2281
										gtt Val			2329
	Ala	Thr	Ala	Cys		Asp	Ile	Ile	Leu	ggc Gly 190	Leu		2377
										gca Ala			2425
										gcc Ala			2473
										ctc Leu			2521 ·

cgg gac											2569
aca caa Thr Gln								_		_	2617
tat cat			Lys T								2665
acc caa Thr Gln 1 290	_			_	-		-				2713
Pro Pro					Суз						2761
ctt tac Leu Tyr			_	_	_		_	_	_	 	2809
ggc gac											2857
aag ggc Lys Gly	_		Pro L	_			_				2905
ggc atc Gly Ile 370		-		_			_		_	 	2953
gac ttt Asp Phe 385			Ser M								3001
ttc acg	Asp Asn		Pro P		Val	Pro			Phe		3049
gcc cat Ala His											3097
gct gcg Ala Ala			Gly T								3145
gtc gcc g Val Ala . 450											3193

atc Ile 465																3241
ccc Pro																3289
tct Ser																3337
gac Asp																3385
acg Thr																3433
tcg Ser 545																3481
act Thr																3529
aag Lys																3577
gag Glu																3625
tac Tyr																3673
gtc Val 625		Ala	Thr	Asp	Ala	Leu	Met		Gly	Phe	Thr	${\tt Gly}$				3721
tca Ser	gtg Val	atc Ile	gac Asp	tgc Cys 645	aat Asn	aca Thr	tgt Cys	gtc Val	açc Thr 650	cag Gln	aca Thr	gtc Val	gac Asp	ttc Phe 655	agc Ser	3769
ctg Leu	gac Asp	ccg Pro	acc Thr 660	ttc Phe	acc Thr	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	gcg Ala	3817
gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	Gly Ggc	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	Gly ggc	3865

	tac Tyr 690															3913
	tcg Ser															3961
	acg Thr															4009
	ej gaa														gtc Val	4057
	aca Thr															4105
	gca Ala 770															4153
	gcc Ala															4201
-	ctc Leu				_		_	_				_		-	_	4249
	agg Arg															4297
	aaa Lys															4345
	acc Thr 850	\mathtt{Trp}			Val		Gly	Val	Leu		Ala					4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	ttc Phe	4489
	gag Glu									cct Pro						4537

Met Gln Leu Ala 915			gca atc ggg Ala Ile Gly 925	
aca gcc acc aag Thr Ala Thr Lys 930				
tgg cgg acc ctc Trp Arg Thr Leu 945	gaa gcc ttc Glu Ala Phe 950	tgg gcg aag Trp Ala Lys	cat atg tgg His Met Trp 955	aat ttc atc 4681 Asn Phe Ile 960
agc ggg ata caa Ser Gly Ile Gln				
gcg ata gca tca Ala Ile Ala Ser 980	Leu Met Ala			
acc acc caa cat Thr Thr Gln His 995	acc ctc ctg Thr Leu Leu	ttt aac atc Phe Asn Ile 1000	ctg ggg gga Leu Gly Gly 1005	Trp Val Ala
gcc caa ctt gct Ala Gln Leu Ala 1010	cct ccc agc Pro Pro Ser 1019	Ala Ala Ser	gct ttc gta Ala Phe Val 1020	ggc gcc ggc 4873 Gly Ala Gly
Ile Ala Gly Ala 1025	Ala Val Gly 1030	agc ata ggc Ser Ile Gly	ctt ggg aag Leu Gly Lys 1035	gtg ctt gtg 4921 Val Leu Val 1040
Ile Ala Gly Ala	Ala Val Gly 1030 ggt tat gga	Ser Ile Gly gca ggg gtg	Leu Gly Lys 1035 gca ggc gcg Ala Gly Ala	Val Leu Val 1040 ctc gtg gcc 4969
Ile Ala Gly Ala 1025 gat att ttg gca	Ala Val Gly 1030 ggt tat gga Gly Tyr Gly 1045 agc ggc gag Ser Gly Glu	gca ggg gtg Ala Gly Val 1050 atg ccc tcc	Leu Gly Lys 1035 gca ggc gcg Ala Gly Ala 0 acc gag gac	Val Leu Val 1040 ctc gtg gcc 4969 Leu Val Ala 1055 ctg gtt aac 5017
Ile Ala Gly Ala 1025 gat att ttg gca Asp Ile Leu Ala ttt aag gtc atg Phe Lys Val Met	Ala Val Gly 1030 ggt tat gga Gly Tyr Gly 1045 agc ggc gag Ser Gly Glu atc ctc tcc Ile Leu Ser	gca ggg gtg Ala Gly Val 1050 atg ccc tcc Met Pro Ser 1065 cct ggc gcc Pro Gly Ala	Leu Gly Lys 1035 gca ggc gcg Ala Gly Ala 0 acc gag gac Thr Glu Asp	Val Leu Val 1040 ctc gtg gcc 4969 Leu Val Ala 1055 ctg gtt aac 5017 Leu Val Asn 1070 ggg gtc gtg 5065 Gly Val Val
Ile Ala Gly Ala 1025 gat att ttg gca Asp Ile Leu Ala ttt aag gtc atg Phe Lys Val Met 1066 cta ctc cct gct Leu Leu Pro Ala	Ala Val Gly 1030 ggt tat gga Gly Tyr Gly 1045 agc ggc gag Ser Gly Glu atc ctc tcc Ile Leu Ser ctg cgt cgg	gca ggg gtg Ala Gly Val 1050 atg ccc tcc Met Pro Ser 1065 cct ggc gcc Pro Gly Ala 1080 cac gtg ggc His Val Gly	Leu Gly Lys 1035 gca ggc gcg Ala Gly Ala 0 acc gag gac Thr Glu Asp cta gtc gtc Leu Val Val 1085 cca ggg gag	Val Leu Val
Ile Ala Gly Ala 1025 gat att ttg gca Asp Ile Leu Ala ttt aag gtc atg Phe Lys Val Met 106 cta ctc cct gct Leu Leu Pro Ala 1075 tgc gca gcg ata Cys Ala Ala Ile	Ala Val Gly 1030 ggt tat gga Gly Tyr Gly 1045 agc ggc gag Ser Gly Glu atc ctc tcc Ile Leu Ser ctg cgt cgg Leu Arg Arg 1095	gca ggg gtg Ala Gly Val 1050 atg ccc tcc Met Pro Ser 1065 cct ggc gcc Pro Gly Ala 1080 cac gtg ggc His Val Gly 5	Leu Gly Lys 1035 gca ggc gcg Ala Gly Ala 0 acc gag gac Thr Glu Asp cta gtc gtc Leu Val Val 1085 cca ggg gag Pro Gly Glu 1100 tcg cgg ggt	Val Leu Val 1040 ctc gtg gcc 4969 Leu Val Ala 1055 ctg gtt aac 5017 Leu Val Asn 1070 ggg gtc gtg 5065 Gly Val Val ggg gct gtg 5113 Gly Ala Val aac cac gtc 5161

cag atc ctc tct Gln Ile Leu Ser 1140	Ser Leu Thr Ile			Leu His
cag tgg atc aac Gln Trp Ile Asn 1155		Thr Pro Cys		
aga gat gtt tgg Arg Asp Val Trp 1170		Thr Val Leu		
tgg ctc cag tcc Trp Leu Gln Ser 1185			Gly Val Pro	
tca tgt caa cgt Ser Cys Gln Arg				
caa acc acc tgc Gln Thr Thr Cys 1220	Pro Cys Gly Ala	_		Lys Asn
tgt tcc atg agg Cys Ser Met Arg 1235		Arg Thr Cys	_	
gga aca ttc ccc Gly Thr Phe Pro 1250		Thr Thr Gly		
ccg gcg cca aat Pro Ala Pro Asn 1265			Val Ala Ala	
tac gtg gag gtt Tyr Val Glu Val				
acc act gac aac Thr Thr Asp Asn 1300		Cys Gln Val	~ ~	Glu Phe
ttc aca gaa gtg Phe Thr Glu Val 1315		Leu His Arg		
aaa ccc ctc cta Lys Pro Leu Leu 1330		Thr Phe Leu		
	2333	•		•

					Leu					cac His					Thr	5929
				Leu					Pro	ccc Pro				Ser	tca Ser	5977
			${\tt Gln}$					Ser		aag Lys			Cys			6025
		Asp					Asp			gag Glu		Asn				6073
	Gln					Asn				gtg Val 1435	Glu					6121
					Ser					caa Gln)					Glu	6169
agg Arg	gaa Glu	gta Val	tcc Ser 1460	Val	ccg Pro	gcg Ala	gag Glu	atc Ile 1465	Leu	cgg Arg	agg Arg	tcc Ser	agg Arg 1470	Lys	ttc Phe	6217
			Met					Arg		gat Asp			Pro			6265
		Ser					Asp			cct Pro		Val				6313
tgt Cys 1509	Pro	ttg Leu	ccg Pro	cct Pro	gcc Ala 1510	Lys	gcc Ala	cct Pro	ccg Pro	ata Ile 1515	Pro	cct Pro	cca Pro	cgg Arg	agg Arg 1520	6361
aag Lys	agg Arg	acg Thr	gtt Val	gtc Val 1525	Leu	tca Ser	gaa Glu	tct Ser	acc Thr 1530	gtg Val)	tct Ser	tct Ser	gcc Ala	ttg Leu 1535	Ala	6409
gag Slu	ctc Leu	gcc Ala	aca Thr 1540	Lys	acc Thr	ttc Phe	Gly	agc Ser 1549	Ser	gaa Glu	tcg Ser	tcg Ser	gcc Ala 1550	Val	gac Asp	6457
agc Ser	ggc	acg Thr 1555	Ala	acg Thr	gcc Ala	tct Ser	cct Pro 1560	Asp	cag Gln	ccc Pro	tcc Ser	gac Asp 1565	Asp	ggc Gly	gac Asp	6505
		Ser					Tyr			atg Met		Pro				6553

	Pro	Gly				Leu					Trp					6601
		gct Ala			Asp					Ser					Trp	6649
		gcc Ala		Ile					Ala					Leu		6697
		gca Ala 163!	Leu					Leu					Leu		tat Tyr	6745
gct Ala	aca Thr 1650	aca Thr	tct Ser	ege Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	6793
gac Asp 1665	Arg	ctg Leu	cag Gln	gtc Val	ctg Leu 1670	Asp	gac Asp	cac His	tac Tyr	cgg Arg 1675	Asp	gtg Val	ctc Leu	aag Lys	gag Glu 1680	6841
		gcg Ala			Ser					Lys					Glu	6889
		tgt Cys		Leu					Ser					Phe		6937
		gca Ala 1715	Lys					Leu					Val			6985
		tcc Ser					Leu					Glu				7033
	Thr	acc Thr				Lys					Cys					7081
		Gly			Pro					Val					Gly	7129
gtt Val	cgt Arg	gtg Val	tgc Cys 1780	Glu	aaa Lys	atg Met	gcc Ala	ctt Leu 1785	Tyr	gat Asp	gtg Val	gtc Val	tcc Ser 1790	Thr	ctc Leu	7177
cct Pro	cag Gln	gcc Ala 1795	Val	atg Met	ggc Gly	tct Ser	tca Ser 1800	Tyr	gga Gly	ttc Phe	caa Gln	tac Tyr 1805	Ser	cct Pro	gga Gly	7225

	gag ttc ctg Glu Phe Leu		Trp Lys Al			73
	gca tat gac Ala Tyr Asp 183	Thr Arg Cys				21
aat gac atc Asn Asp Ile	cgt gtt gag Arg Val Glu 1845	gag tca atc Glu Ser Ile	tac caa tg Tyr Gln Cy 1850	t tgt gac s Cys Asp	ttg gcc 736 Leu Ala 1855	59
	aga cag gcc Arg Gln Ala 1860		Leu Thr Gl		Tyr Ile	17
	ctg act aat Leu Thr Asn 5					5 5
	agc ggt gta Ser Gly Val		Ser Cys Gl			L3
	aag gcc gct Lys Ala Ala 191	Ala Ala Cys				51
	ctc gta tgc Leu Val Cys 1925					9
	caa gag gac Gln Glu Asp 1940		Leu Arg Al		Glu Ala	57
	tac tct gcc Tyr Ser Ala 5)5
gac ttg gag Asp Leu Glu 1970	ttg ata aca Leu Ile Thr	tca tgc tcc Ser Cys Ser 1975	tcc aat gt Ser Asn Va 19	l Ser Val	gcg cac 775 Ala His	53
gat gca tct Asp Ala Ser 1985	ggc aaa agg Gly Lys Arg 1990	Val Tyr Tyr	ctc acc cg Leu Thr Ar 1995	t gac ccc g Asp Pro	acc acc 780 Thr Thr 2000)1
ccc ctt gcg Pro Leu Ala	cgg gct gcg Arg Ala Ala 2005	tgg gag aca Trp Glu Thr	gct aga ca Ala Arg Hi 2010	c act cca s Thr Pro	gtc aat 784 Val Asn 2015	19
tcc tgg cta Ser Trp Leu	ggc aac atc Gly Asn Ile 2020	atc atg tat Ile Met Tyr 202	Ala Pro Th	c ttg tgg r Leu Trp 2030	Ala Arg	∍7

atg atc ctg atg a Met Ile Leu Met 7 2035			_		7945
ctt gaa aaa gcc o Leu Glu Lys Ala I 2050		Gln Ile Ty			7993
gag cca ctt gac o Glu Pro Leu Asp I 2065					8041
gca ttt tca ctc o Ala Phe Ser Leu P			Glu Ile Asn A		8089
tca tgc ctc agg a Ser Cys Leu Arg I 2100			Leu Arg Val T		8137
cgg gcc aga agt g Arg Ala Arg Ser V 2115					8185
gcc act tgt ggc a Ala Thr Cys Gly I 2130	aag tac ctc Lys Tyr Leu 2139	Phe Asn Tr	g gca gta agg ac o Ala Val Arg Tl 2140	cc aag ctc ir Lys Leu	8233
aaa ctc act cca a Lys Leu Thr Pro 1 2145	atc ccg gct Ile Pro Ala 2150	gcg tcc cag Ala Ser Glr	g ttg gat tta to 1 Leu Asp Leu So 2155	ec agc tgg er Ser Trp 2160	8281
ttc gtt gct ggt t Phe Val Ala Gly 7			Tyr His Ser Le		8329
gcc cga ccc cgc t Ala Arg Pro Arg T 2180			ı Leu Leu Leu Se		8377
gta ggc atc tat c Val Gly Ile Tyr I 2195	Leu Leu Pro		a acggggagct aaa	acactcca	8427
ggccaatagg ccatco ttttttttt tttttt tttggtggct ccatct actgcagaga gtgctg	tttt ttcccc	ccct tttttc cacg gctago	ecct tttttttco	ttttctttcc	8547
<210> 6 <211> 8638 <212> DNA <213> HCV <220> <221> CDS <222> (1802)(8	3407)				
· <400> 6					

accageco															
tcttcacg	gca g	gaaag	gegt	et ag	gccai	rggcg	g tta	agta	tgag	tgt	gtg	cag	cctc	caggac	120
ccccctc															
gacgaccg															
gcgagact															
gtgcttgc															
ctcaaaga	laa a	lacca	aagg	ag c	gcgc	catga	a tt	gaac	aaga	tgga	attg	cac	gcag	gttete	420
cggccgct															
ctgatgcc															
acctgtcc															
cgacgggc															
tgctattg															
aagtatco	eat c	atgg	getga	it go	caat	gcggd	gg	etge	atac	gcti	tgat	ccg	gcta	cctgcc	780
cattcgac	ca c	caag	gcgaa	aa ca	atcg	catco	gago	cgag	cacg	tact	tegg	atg	gaag	ccggtc	840
ttgtcgat	ca g	gate	jatci	gg	acga	agago	c ato	cagg	ggct	cgc	gcca	acc i	gaact	tgttcg	900
ccaggete															
gcttgccg															
tgggtgtg	igc g	gaco	gcta	at ca	agga	catag	g cgt	ttgg	ctac	ccgt	tgata	att 9	gctga	aagagc	1080
ttggcggc															
agcgcatc															
cagaccac															
ccgaagcc															
gccgtctt	tt g	gcaa	tgtg	ga gg	ggcc	cggaa	a aco	ctgg	ccct	gtc	tcti	tga (cgago	cattcc	1380
taggggtc															
agttecte															
gaaccccc															
tacaaaaa		rcaca	lacco	cc ac	ataca	cacgt	tgt:	tgagi	ttgg	atag	gttgi	tgg a	aaaga	agtcaa	1620
Lycadayy	ica g	,											_		
atggctct	cc t	caag	jcgta	it to	caaca	aaggg	g gct	tgaag	ggat	gcc	cagaa	agg 1	tacco	ccattg	1680
atggctct tatgggat	cc t	caag	gggg	it to	caaca toggt	aaggg tgċac	ate	gctti	taca	tgt	caga: gttt:	agt (tacco cgago	ccattg gttaaa	1740
atggctct tatgggat aaacgtct	cc t ct g ag g	caag atct cccc	ccga gggg	it to jc ct ia co	caaca ceggi caegg	aaggg tgċac gggac	ate gte	gctti ggtti	taca ttcc	tgtg	caga: yttt: yaaa:	agt (tacco cgago acgat	ccattg gttaaa taatac	1740 1800
atggctct tatgggat aaacgtct c atg ga	cc t ct g ag g c cg	caag atct cccc g ga	gggg gggg ccga g at	t to gc ct ta co	caaca teggi caegg ca go	aaggg tgċad gggad ca td	ate gte	getti ggtti ge gg	taca ttcc ga gg	tgtg tttg gc gc	cagaa gttta gaaaa cg gi	agt d aac a tt't!	tacco cgago acgat tc gt	ccattg gttaaa taatac ta ggt	1740 1800
atggctct tatgggat aaacgtct c atg ga Met As	cc t ct g ag g c cg	caag atct cccc g ga	gggggggggggggggggggggggggggggggggggggg	it to go of la co eg go et Al	caaca teggi caegg ca go	aaggg tgċad gggad ca td	ate gte	getti getti ge gg ys G	taca ttcc ga gg ly G]	tgtg tttg gc gc	cagaa gttta gaaaa cg gi	agt d aac a tt't!	tacco cgago acgat tc gt he Va	ccattg gttaaa taatac ta ggt al Gly	1740 1800
atggctct tatgggat aaacgtct c atg ga	cc t ct g ag g c cg	caag atct cccc g ga	gggg gggg ccga g at	it to go of la co eg go et Al	caaca teggi caegg ca go	aaggg tgċad gggad ca td	ate gte	getti getti ge gg ys G	taca ttcc ga gg	tgtg tttg gc gc	cagaa gttta gaaaa cg gi	agt d aac a tt't!	tacco cgago acgat tc gt he Va	ccattg gttaaa taatac ta ggt	1740 1800
atggctct tatgggat aaacgtct c atg ga Met As	cc t ct g ag g c cg	caag atct geec g ga g Gl	gggta gggg ccga g at u Me	it to je et ia eo ig go it Al	caaca teggt caegg ca go la Al	aaggg tgċac gggac ca tc la Se	e atg e gtg eg tg er Cy	gctti ggtti gc gg ys G	taca ttcc ga gg ly G] 10	tgtg tttg gc gc ly Al	cagaa yttta yaaaa cg gi la Va	agt d aac a tt'ti al Pl	tacco cgagg acgat tc gt he Va	ccattg gttaaa taatac ta ggt al Gly	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata	ctc tct gag gag acc cg	caag atct cccc g ga g Gl	gggggecega g at u Me	ttg	caaca teggt caegg ca go la Al	aaggg tgċac gggac ca tc la Se	cac	getti getti ge gg ys G: :	taca ttee ga gg ly Gl 10	tgtg tttg gc gc ly Al	cagaa gttta gaaaa cg gi la Va	agt of all Pl	tacco cgago acgat tc gt he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1740 1800
atggctct tatgggat aaacgtct c atg ga Met As	ctc tct gag gag acc cg	ttg Leu	gggggecega g at u Me	ttg	caaca teggt caegg ca go la Al	aaggg tgċac gggac ca tc la Se	cac His	getti getti ge gg ys G: :	taca ttee ga gg ly Gl 10	tgtg tttg gc gc ly Al	cagaa gttta gaaaa cg gi la Va	agt (ac a tt to al P) ctc Leu	tacco cgagg acgat tc gt he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata	ctc tct gag gag acc cg	caag atct cccc g ga g Gl	gggggecega g at u Me	ttg	caaca teggt caegg ca go la Al	aaggg tgċac gggac ca tc la Se	cac	getti getti ge gg ys G: :	taca ttee ga gg ly Gl 10	tgtg tttg gc gc ly Al	cagaa gttta gaaaa cg gi la Va	agt of all Pl	tacco cgagg acgat tc gt he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile	cc t ct g ag g c cg p Ar ctc Leu	ttg Leu	gegta gggg eccga g at u Ma s acc Thr	at to go of aa co gg go et Al ; ttg Leu	caaca teggt caegg ca ge la Al tea Ser	aaggg tgèac gggac ca to la Se ceg Pro	c atomic gto get Cycles Cycles Cycles Cycles Cycles Cycles Cycles Cycles Cycles Cac His 25	getti ggtti ge gg ys G ; tat Tyr	taca ttcc ga gg ly Gl 10 aag Lys	tgtg tttg gc gc ly Al ctg Leu	cagaa gttta gaaaa eg gt la Va ttc Phe	agt (aac a tt tr al P ctc Leu 30	tacco cgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile	cc t cag g c cg p Ar ctc Leu	ttg ttgg	gegta gggg ceega g at u Me acc Thr	t to ge et aa ec g ge et Al ; ttg Leu	caaca teggt caegg ca ge la Al tea Ser	aaggg tgéac gggac ca to la Se ccg Pro	c ato	getti getti ge gg ys G tat Tyr	taca ttcc ga gg ly Gl 10 aag Lys	tgtg tttg gc gc ly Al ctg Leu gcc	cagas gttts gaaas cg gt la Va ttc Phe	agt aac at the state of the sta	taccoccgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile	cc t cct g cag g cc cg p Ar ctc Leu tgg	ttg ttgg	gegta gggg ceega g at u Me acc Thr	t to ge et aa ec g ge et Al ; ttg Leu	caaca teggt caegg ca ge la Al tea Ser	aaggg tgéac gggac ca to la Se ccg Pro ttt Phe	c ato	getti getti ge gg ys G tat Tyr	taca ttcc ga gg ly Gl 10 aag Lys	tgtg tttg gc gc ly Al ctg Leu gcc	gagaggaggaggaggaggaggaggaggaggaggaggagg	agt aac at the state of the sta	taccoccgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile	cc t cag g c cg p Ar ctc Leu	ttg ttgg	gegta gggg ceega g at u Me acc Thr	t to ge et aa ec g ge et Al ; ttg Leu	caaca teggt caegg ca ge la Al tea Ser	aaggg tgéac gggac ca to la Se ccg Pro	c ato	getti getti ge gg ys G tat Tyr	taca ttcc ga gg ly Gl 10 aag Lys	tgtg tttg gc gc ly Al ctg Leu gcc	cagas gttts gaaas cg gt la Va ttc Phe	agt aac at the state of the sta	taccoccgagg acgat tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile	cc t ag g ac cg p Ar ctc Leu tgg Trp 35	ttg Leu 20	gegta gggg eeega g at u Me acc Thr tta Leu	ttg caa Gln	tca tcaggla Al tca Ser tat Tyr	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40	cac His 25 atc	getti ggtti ge gg ys G tat Tyr acc Thr	taca ttcc ga gg ly GJ 10 aag Lys agg Arg	tgtg tttg gc gc ly Al ctg Leu gcc Ala	ttc pagg gttta gagas gg gt la Va ttc Phe gag Glu 45	egt (ac a at train p) ctc Leu 30 gca Ala	taccoccgaggacgathe Value Got Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1740 1800 1849 1897
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile ctc ata	cc t ag g c cg p Ar ctc Leu tgg Trp 35	ttg Leu 20 tgg tag	gegta gggg eeega g at u Me acc Thr tta Leu	ttg caa caa Gln	tca tcaggi cacggi ca go la Al tca Ser tat Tyr	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40 aac	cac cac His 25 atc Ile	getti ggtti ge gg ys G tat Tyr acc Thr	taca ttcc ga gg ly Gl 10 aag Lys agg Arg	tgtg ttttg gc gg ly Al ctg Leu gcc Ala	ttc pagg tttc gag ttc Phe gag Glu 45	egt dack at the state of the st	taccoccgaggacgatccgatccgatccgatccgatccga	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1740 1800 1849
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val	cc t ag g c cg p Ar ctc Leu tgg Trp 35	ttg Leu 20 tgg tag	gegta gggg eeega g at u Me acc Thr tta Leu	ttg caa caa Gln	tca tca ca go la Al tca Ser tat Tyr	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40 aac	cac cac His 25 atc Ile	getti ggtti ge gg ys G tat Tyr acc Thr	taca ttcc ga gg ly Gl 10 aag Lys agg Arg	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc Gly	ttc pagg tttc gag ttc Phe gag Glu 45	egt dack at the state of the st	taccoccgaggacgatccgatccgatccgatccgatccga	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1740 1800 1849 1897
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile ctc ata	cc t ag g c cg p Ar ctc Leu tgg Trp 35	ttg Leu 20 tgg tag	gegta gggg eeega g at u Me acc Thr tta Leu	ttg caa caa Gln	tca tcaggi cacggi ca go la Al tca Ser tat Tyr	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40 aac	cac cac His 25 atc Ile	getti ggtti ge gg ys G tat Tyr acc Thr	taca ttcc ga gg ly Gl 10 aag Lys agg Arg	tgtg ttttg gc gg ly Al ctg Leu gcc Ala	ttc pagg tttc gag ttc Phe gag Glu 45	egt dack at the state of the st	taccoccgaggacgatccgatccgatccgatccgatccga	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1740 1800 1849 1897
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50	cc t ct g cc cg c Ar ctc Leu tgg Trp 35	ttg teg teg ttg teu 20 tgg Trp	gegta gggg cega g at u Me acc Thr tta Leu	ttg caa Gln ccc Pro	tca tca ca go la Al tca Ser tat Tyr ctc Leu 55	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40 aac Asn	cate gte Cy cac His 25 atc Ile ytt Val	getti getti ge gg ys Gi tat Tyr acc Thr	taca ttcc ga gg ly Gl 10 aag Lys agg Arg ggg Gly	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60	ttc paag tttc phe gag Glu 45 cgc	ctc Leu 30 gca Ala gat	taccoccgaggacgate Value	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1740 1800 1849 1897 1945
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50 atc ctc	cc t ct g cc cg c cg ctc Leu tgg Trp 35 tgg Trp ctc	ttg Leu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro	ttg caa Gln ccc Pro	tca tca ca go la Al tca Ser tat Tyr ctc Leu 55	cac	c atg c gtg cg tg cac His 25 atc Ile gtt Val	getti getti ge gg ys G tat Tyr acc Thr cgg Arg	taca ttcc ga gg ly Gl 10 aag Lys agg Arg Gly cta	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60 atc	ttc gaag tttc gaaa gg gt la Va ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccgaggacgattc gthe Vala cac His	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1740 1800 1849 1897
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50 atc ctc Ile Leu	cc t ct g cc cg c cg ctc Leu tgg Trp 35 tgg Trp ctc	ttg Leu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro	ttg caa caa Gln ccc pro	tca ser tat tat Tyr ctc Leu 55	cac	c atg c gtg cg tg cac His 25 atc Ile gtt Val	getti getti ge gg ys G tat Tyr acc Thr cgg Arg	taca ttcc ga gg ly Gl 10 aag Lys agg Arg ggg Gly cta Leu	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60 atc	ttc gaag tttc gaaa gg gt la Va ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccagage acgate to get Ala cac His	ccattg gttaaa caatac ca ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr	1740 1800 1849 1897 1945
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50 atc ctc	cc t ct g cc cg c cg ctc Leu tgg Trp 35 tgg Trp ctc	ttg Leu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro	ttg caa Gln ccc Pro	tca ser tat tat Tyr ctc Leu 55	cac	c atg c gtg cg tg cac His 25 atc Ile gtt Val	getti getti ge gg ys G tat Tyr acc Thr cgg Arg	taca ttcc ga gg ly Gl 10 aag Lys agg Arg Gly cta	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60 atc	ttc gaag tttc gaaa gg gt la Va ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccagage acgate to get Ala cac His	gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1740 1800 1849 1897 1945
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50 atc ctc Ile Leu 65	cc t ct g ac cg p Ar ctc Leu tgg Trp 35 Trp ctc Leu	ttg tecu tggg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ecc Pro	ttg caa Gln ccc Pro gcg Ala	tca tca gela Al tca Ser tat Tyr ctc Leu 55 atc	cac His	c atg c gtg cg tg cac His 25 atc Ile gtt Val cca	tat Tyr acc Thr cgg Glu	taca ttcc ga gg ly Gl 10 aag Lys agg Arg ggg Gly cta Leu 75	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile	ttc gag tttc gag ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1740 1800 1849 1897 1945 1993
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ata Leu Ile cta ctc ata Leu Ile caa gtg Gln Val 50 atc ctc Ile Leu 65 aaa atc	cc t g g g c c c c L e u t g g T r p c t c L e u t t g g T r p c t c L e u t t g	ttg teu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro gcg Ala 70 ata	tca ser tca ser tat Tyr ctc Leu 55 atc Ile	cac His	c atg c gtg cg tg cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Glu ctc	taca ttcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile	ttc gag tttc gag ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile caa gtg Gln Val 50 atc ctc Ile Leu 65	cc t g g g c c c c L e u t g g T r p c t c L e u t t g g T r p c t c L e u t t g	ttg teu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro gcg Ala 70 ata	tca ser tca ser tat Tyr ctc Leu 55 atc Ile	cac His	c atg c gtg cg tg cac His 25 atc Ile gtt Val cca Pro	gettingttinge gettinge getting getting gettinge getting getting getting getting getting getting getting getting	taca ttcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile	ttc gag tttc gag ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945 1993
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ata Leu Ile cta ctc ata Leu Ile caa gtg Gln Val 50 atc ctc Ile Leu 65 aaa atc	cc t g g g c c c c L e u t g g T r p c t c L e u t t g g T r p c t c L e u t t g	ttg teu 20 tgg Trp atc Ile	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro gcg Ala 70 ata	tca ser tca ser tat Tyr ctc Leu 55 atc Ile	cac His	c atg c gtg cg tg cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Glu ctc	taca ttcc ga gg ly Gl l0 aag Lys agg Arg ggg Gly cta Leu 75 atg	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile	ttc gag tttc gag ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1740 1800 1849 1897 1945 1993
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ggg Gln Val 50 atc ctc Ile Leu 65 aaa atc Lys Ile	cc t cag g cp Ar ctc Leu tgg Trp ctc Leu ttgg	ttg tgg Gl ttgg Trp atc Ile acg	gegta gggg cega g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg cet Alicia caa Gln ccc Pro gcg Ala 70 ata Ile	tca caeggina Al tca ser tat Tyr ctc Leu 55 atc Ile ctc	cac His	c atgcg tgg tgg tgg tgg tgg tgg tgg tgg tgg	tat Tyr acc Thr cgg Glu ctc Leu 90	taca ttcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75 atg Met	tgtg ttttg gc gd ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc phe gag ctt gaaa cg gt la Va ttc phe gag ctt Arg ttt phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile gct Ala 95	ttg Leu gtc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993 2041
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ata ata atc ata acc	cc t g g cg cp Ar ctc Leu tgg Trp ctc Leu ttg Leu ttg aaa	ttg theu tgg Trp atc Ile acg Thr	gegta geggg cegga g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tca ser tat Tyr ctc Leu 55 atc Ile ctc	cac His	c atgcg tgg tgg tgg tgg tgg tgg tgg tgg tgg	tat Tyr acc Thr cgg Glu ctc Leu 90 gca	taca ttcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75 atg Met cac	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc pag tttc gag ttc Phe gag Glu 45 cgc Arg ttt Phe ctc	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	gct Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ggg Gln Val 50 atc ctc Ile Leu 65 aaa atc Lys Ile	cc t g g cg cg p Ar ctc Leu tgg Trp ctc Leu ttgu aaa Lys	ttg ttgu tgg Gl ttgu tgg Trp atc Leu gtg Val	gegta geggg cegga g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tca ser tat Tyr ctc Leu 55 atc Ile ctc	cac His	cated gto Cac His 25 atc Ile gtt Val cca Pro cgc Arg	tat Tyr acc Thr cgg Glu ctc Leu 90 gca	taca ttcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75 atg Met cac	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc pag tttc gag ttc Phe gag Glu 45 cgc Arg ttt Phe ctc	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att Ile	gct Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993 2041
atggctct tatgggat aaacgtct c atg ga Met As 1 ctg ata Leu Ile ctc ata Leu Ile cta ata ata atc ata acc	cc t g g cg cg p Ar ctc Leu tgg Trp ctc Leu ttgu aaa Lys	ttg theu tgg Trp atc Ile acg Thr	gegta geggg cegga g at u Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc Pro gcg Ala 70 ata Ile	tca ser tca ser tat Tyr ctc Leu 55 atc Ile ctc	cac His	c atgcg tgg tgg tgg tgg tgg tgg tgg tgg tgg	tat Tyr acc Thr cgg Glu ctc Leu 90 gca	taca ttcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75 atg Met cac	tgtg ttttg gc gc ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc pag tttc gag ttc Phe gag Glu 45 cgc Arg ttt Phe ctc	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	gct Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1740 1800 1849 1897 1945 1993 2041

	Met			cgg Arg												2185
				gcc Ala												2233
				gac Asp												2281
				gtc Val 165												2329
				acc Thr	_		_		-			_		_		2377
				agg Arg												2425
				gly aaa		-						-	_			2473
				ggc Gly												2521
				cag Gln 245												2569
				ctg Leu												2617
				Gly												2665
				acc Thr												2713
				cgt Arg												2761
ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	agg Arg	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	cgg Arg	2809

						cta Leu										2857
aag Lys	ggc	tct Ser 355	tcg Ser	ggc Gly	ggt Gly	cca Pro	ctg Leu 360	ctc Leu	tgc Cys	ccc Pro	tcg Ser	365 365	cac His	gct Ala	gtg Val	2905
Gly	atc Ile 370	ttt Phe	cgg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cga Arg	Gly 333	gtt Val 380	gcg Ala	aag Lys	gcg Ala	gtg Val	2953
gac Asp 385	ttt Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gga Gly	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	gtc Val 400	3001
						cct Pro										3049
gcc Ala	cat His	cta Leu	cac His 420	gcc Ala	cct Pro	act Thr	ggt Gly	agc Ser 425	ggc Gly	aag Lys	agc Ser	act Thr	aag Lys 430	gtg Val	ccg Pro	3097
gct Ala	gcg Ala	tat Tyr 435	gca Ala	gcc Ala	caa Gln	ej aaa	tat Tyr 440	aag Lys	gtg Val	ctt Leu	gtc Val	ctg Leu 445	aac Asn	.ccg Pro	tcc Ser	3145
gtc Val	gcc Ala 450	gcc Ala	acc Thr	cta Leu	ggt Gly	ttc Phe 455	gly gag	gcg Ala	tat Tyr	atg Met	tct Ser 460	aag Lys	gca Ala	cat His	ggt Gly	3193
atc Ile 465	gac Asp	cct Pro	aac Asn	atc Ile	aga Arg 470	acc Thr	gly aaa	gta Val	agg Arg	acc Thr 475	atc Ile	acc Thr	acg Thr	ggt Gly	gcc Ala 480	3241
ccc Pro	atc Ile	acg Thr	tac Tyr	tcc Ser 485	acc Thr	tat Tyr	ggc Gly	aag Lys	ttt Phe 490	ctt Leu	gcc Ala	gac Asp	ggt Gly	ggt Gly 495	tgc Cys	3289
tct Ser	gly ggg	ggc Gly	gcc Ala 500	tat Tyr	Asp	atc Ile	Ile	Ile	Cys	Asp	gag Glu	Cys	His	Ser	act Thr	3337
gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc	atc Ile 520	ggc Gly	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
acg Thr	gct Ala 530	gga Gly	gcg Ala	cga Arg	ctc Leu	gtc Val 535	gtg Val	ctc Leu	gcc Ala	acc Thr	gct Ala 540	acg Thr	cct Pro	ccg Pro	gga Gly	3433
tcg Ser 545	gtc Val	acc Thr	gtg Val	cca Pro	cat His 550	cca Pro	aac Asn	atc Ile	gag Glu	gag Glu 555	gtg Val	gct Ala	ctg Leu	tcc Ser	agc Ser 560	3481

										atc Ile						3529
_								_		tcc Ser	_	_		_	_	3577
		_		_	_					ċtc Leu		_	_			3625
				_	-		_			act Thr	_		_	_		3673
										ttt Phe 635						3721
										cag Gln						3769
_	_	_					-		_	acc Thr				_	~ ~	3817
			_			_				ggt Gly				_		3865
								_		ccc Pro	_		_		_	3913
										ggc Gly 715						3961
					Thr		Val	Arg		cgg Arg					Thr	4009
										gag Glu						4057
							_	_		ttc Phe	_		_		_	4105
										gca Ala						4153

						cca Pro										4201
						cct Pro										4249
						caa Gln										4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
						ggc Gly 855										4393
						gtg Val								Leu		4441
						ccc Pro										4489
						gcc Ala										4537
						ttc Phe									caa Gln	4585
aca Thr	gcc Ala 930	acc Thr	aag Lys	caa Gln	gcg Ala	gag Glu 935	gct Ala	gct Ala	gct Ala	ccc Pro	gtg Val 940	gtg Val	gaa Glu	tcc Ser	aag Lys	4633
						ttc Phe		Ala	Lys		Met					4681
						gca Ala										4729
gcg Ala	ata Ile	gca Ala	tca Ser 980	ctg Leu	atg Met	gca Ala	ttc Phe	aca Thr 985	gcc Ala	tct Ser	atc Ile	acc Thr	agc Ser 990	ccg Pro	ctc Leu	4777
			His			ctg Leu		Asn					Trp			4825

								54 /	93							
		Leu					Ala					Val	Gly			4873
	Ala					Gly					Gly		gtg Val			4921
					Tyr					Ala			ctc Leu		Ala	4969
	_	_	_	Ser			_		Ser			_	ctg Leu 107	Val		5017
			Ala					Gly	_		_	_	g GJÀ aaa	_		5065
	_	Ala		_	_		His					Glu	ely aaa	_		5113
_	Trp	_			_	Ile			_		Arg		aac Asn		_	5161
					Val					Ala			cgt Arg		Thr	5209
				Ser					Gln				agg Arg 115	Leu		5257
			Asn					Thr					tcg Ser 5			5305
		Val					Cys					Asp	ttc Phe			5353
	Leu					Leu					Gly		ccc Pro			5401
					Tyr					Arg			ggc Gly		Met	5449
				Pro					Ile				gtg Val 123	Lys		5497

			Arg					Arg			agt Ser		Thr			5545
		Phe					Tyr				ccc Pro 1260	Cys				5593
	Ala					Arg					gtg Val					5641
					Arg					His	tac Tyr				Met	5689
				Val					Gln		ccg Pro			Glu		5737
			Val					Leu			tac Tyr		Pro			5785
		Leu					Val				gtc Val 1340	Gly				5833
	Leu					Leu					gaa Glu					5881
					Leu					His	att Ile				Thr	5929
				Leu					Pro		tcc Ser			Ser		5977
			Gln					Ser			gca Ala		Cys			6025
cgt Arg	cat His 1410	Asp	tcc Ser	ccg Pro	gac Asp	gct Ala 1415	Asp	ctc Leu	atc Ile	gag Glu	gcc Ala 1420	Asn	ctc Leu	ctg Leu	tgg Trp	6073
egg Arg 1425	${\tt Gln}$	gag Glu	atg Met	ggc	999 Gly 1430	Asn	atc Ile	acc Thr	cgc Arg	gtg Val 1435	gag Glu	tca Ser	gaa Glu	aat Asn	aag Lys 1440	6121
gta Val	gta Val	att Ile	ttg Leu	gac Asp 1445	Ser	ttc Phe	gag Glu	ccg Pro	ctc Leu 1450	Gln	gcg Ala	gag Glu	gag Glu	gat Asp 1455	Glu	6169

Arg Glu Val			g cgg agg tcc u Arg Arg Ser		
			g gat tac aac o Asp Tyr Asn 148	Pro Pro	
	Trp Lys Asp P		c cct cca gtg l Pro Pro Val 1500		
			g ata cca cct o Ile Pro Pro 1515		
			c gtg tct tct r Val Ser Ser 30		Ala
Glu Leu Ala			c gaa tcg tcg r Glu Ser Ser		
			g ccc tcc gac n Pro Ser Asp 156	Asp Gly	
	Asp Val Glu S	_	c atg ccc ccc r Met Pro Pro 1580		
			g cct tgg tct y Pro Trp Ser 1595		
		al Val Cys Cy	c tcg atg tcc s Ser Met Ser 10		Trp
Thr Gly Ala	Leu Ile Thr F	ro Cys Ala Al	g gag gaa acc a Glu Glu Thr	Lys Leu	
	Leu Ser Asn S		t cac cac aac g His His Asn 164	Leu Val	
	Ser Arg Ser A		g cag aag aag g Gln Lys Lys 1660		

atg aag gcg aag go Met Lys Ala Lys Al 10				
gaa gcc tgt aag ct Glu Ala Cys Lys Le 1700	eu Thr Pro Pro B			Phe Gly
tat ggg gca aag ga Tyr Gly Ala Lys As 1715				
atc cgc tcc gtg to Ile Arg Ser Val Ti 1730		eu Glu Asp T		
gac acc acc atc at Asp Thr Thr Ile Me 1745	tg gca aaa aat g et Ala Lys Asn G 1750	gag gtt ttc t lu Val Phe C 1755	gc gtc caa Tys Val Gln	cca gag .7081 Pro Glu 1760
	ys Pro Ala Arg I 765	eu Ile Val F 1770	Phe Pro Asp	Leu Gly 1775
gtt cgt gtg tgc ga Val Arg Val Cys Gl 1780	lu Lys Met Ala I 1	eu Tyr Asp V .785	Val Val Ser 1790	Thr Leu
cct cag gcc gtg at Pro Gln Ala Val Me 1795				
cag cgg gtc gag tt Gln Arg Val Glu Ph 1810	tc ctg gtg aat g ne Leu Val Asn A 1815	la Trp Lys A	gcg aag aaa lla Lys Lys 1820	tgc cct 7273 Cys Pro
atg ggc ttc gca ta Met Gly Phe Ala Ty 1825	at gac acc cgc t yr Asp Thr Arg C 1830	gt ttt gac t ys Phe Asp S 1835	ca acg gtc Ser Thr Val	act gag 7321 Thr Glu 1840
aat gac atc cgt gt Asn Asp Ile Arg Va 18		le Tyr Gln C	Cys Cys Asp	
ccc gaa gcc aga ca Pro Glu Ala Arg Gl 1860	ln Ala Ile Arg S	cg ctc aca g er Leu Thr G .865	gag cgg ctt Blu Arg Leu 1870	Tyr Ile
ggg ggc ccc ctg ac Gly Gly Pro Leu Th 1875	ct aat tot aaa g ar Asn Ser Lys G 1880	gg cag aac t ly Gln Asn C	gc ggc tat Cys Gly Tyr 1885	cgc cgg 7465 Arg Arg
tgc cgc gcg agc gg Cys Arg Ala Ser Gl	rt ata eta aca a	ec age toe o	or aat acc	ctc aca 7513

tgt tac ttg aag g Cys Tyr Leu Lys 1 1905		Ala Cys Arg A			7561
tgc acg atg ctc g Cys Thr Met Leu 1					7609
gcg ggg acc caa g Ala Gly Thr Gln (1940	Glu Asp Glu A		rg Ala Phe T		7657
atg act aga tac t Met Thr Arg Tyr S 1955	Ser Ala Pro P		_	_	7705
gac ttg gag ttg a Asp Leu Glu Leu 1 1970					7753
gat gca tct ggc a Asp Ala Ser Gly 1 1985		Tyr Tyr Leu T			7801
ccc ctt gcg cgg g Pro Leu Ala Arg A					7849
tcc tgg cta ggc a Ser Trp Leu Gly 2 2020	Asn Ile Ile M		ro Thr Leu I		7897
atg atc ctg atg a Met Ile Leu Met 3 2035	Thr His Phe P				7945
ctt gaa aaa gcc o Leu Glu Lys Ala I 2050					7993
gag cca ctt gac o Glu Pro Leu Asp 1 2065		le Ile Gln A	rg Leu His G	ly Leu Ser	8041
gca ttt tca ctc c Ala Phe Ser Leu I					8089
•	2085	2090	iu ile Ash A	2095	
tca tgc ctc agg a Ser Cys Leu Arg I 2100	2085 aaa ctt ggg g Lys Leu Gly V	2090 Ita eeg eee ti	tg cga gtc t eu Arg Val T	2095 :gg aga cat	8137

gcc act tgt Ala Thr Cys 2130		s Tyr :					Arg				8233
aaa ctc act Lys Leu Thr 2145			Ala Ala			Asp					8281
ttc gtt gct Phe Val Ala		r Ser			e Tyr					Arg	8329
gcc cga ccc Ala Arg Pro									Val		8377
gta ggc atc Val Gly Ile 219	Tyr Le			Arg *	a acg	gggag	rct a	aac	actco	ca ·	8427
ggccaatagg ttttttttt ccatcttagc gtgctgatac	ttttttt cctagtc	ttt tt acg gc	ttttcctc tagctgtg	ttttt aaagg	tttcc	tttt	cttt	cc	tttgg	gtggct	8547
<210> 7 <211> 8638 <212> DNA <213> HCV											
<220> <221> CDS <222> (1802) (84	07)	١								
<400> 7											
gccagccccc		_		_			_	_		_	
tcttcacgca											
ccccctccc											
gacgaccggg			_		_					-	
gcgagactgc gtgcttgcga		_				-			_		
ctcaaagaaa							_	_			
cggccgcttg											
ctgatgccgc											
acctgtccgg											
cgacggcgt											
tgctattggg											
aagtatccat											
cattcgacca											
ttgtcgatca											
ccaggeteaa											
gcttgccgaa tgggtgtggc											
-999191990	ggaccyc										
	atoggget	מאכ ככ	CEECCECC		T 2 ~~~	Par-		יייד		755000	7740
ttggcggcga											
ttggcggcga agcgcatcgc	cttctate	cgc ct	tcttgacg	, agttc	ttctg	agtt	cgcg	gcc	caga	tgttaa	1200
ttggcggcga agcgcatcgc cagaccacaa	cttctate cggtttc	ege et eet et	tcttgacg agcgggat	, agttc : caatt	ttetg cegee	agtt cccc	cgcc	gcc cta	caga: acgt:	tgttaa tactgg	1200 1260
ttggcggcga agcgcatcgc	cttctat cggtttc ttggaat	cgc ct cct ct aag gc	tcttgacg agcgggat cggtgtgc	g agttc caatt gtttg	ttetg cegce tetat	agtt cccc atgt	cgcg cccc tatt	gcc cta ctt	caga: acgt: ccac	tgttaa tactgg catatt	1200 1260 1320

taggggtett tececteteg ccaaaggaat gcaaggtetg ttgaatgteg tgaaggaage 1440 agtteetetg gaagettett gaagacaaac aacgtetgta gcgaccettt gcaggcageg 1500 gaacececca cetggcgaca ggtgcetetg cggccaaaag ccacgtgtat aagatacace 1560 tgcaaaggeg gcacaacece agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620 atggetetee tcaagegtat tcaacaaggg gctgaaggat gcccagaagg taceccattg 1680 tatgggatet gatetgggge ctcggtgcac atgetttaca tgtgtttagt cgaggttaaa 1740 aaacgtetag gcccccgaa ccacggggac gtggttttee tttgaaaaac acgataatac 1800 c atg gac cgg gag atg gca gca tcg tgc gga gge gcg gtt tte gta ggt 1849 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1 5 10 15	
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg 1897 Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg 20 25 30	,
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu 35 40 45	;
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc 1993 Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val 50 55 60	ļ
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr 65 70 75 . 80	L
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt 2089 Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly 85 90 95)
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca 2137 Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala 100 105 110	7
tgc atg ctg gtg cgg aag gtt gct ggg ggt cat tat gtc caa atg gct 2185 Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala 115 120 125	5
ctc atg aag ttg gcc gca ctg aca ggt acg tac gtt tat gac cat ctc 2233 Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu 130 135 140	3
acc cca ctg cgg gac tgg gcc cac gcg ggc cta cga gac ctt gcg gtg Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val 145 150 155 160	L
gca gtt gag ccc gtc gtc ttc tct gat atg gag acc aag gtt atc acc 2329 Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr 165 170 175	€
tgg ggg gca gac acc gcg gcg tgt ggg gac atc atc ttg ggc ctg ccc 2377 Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro 180 185 190	7
gtc tcc gcc cgc agg ggg agg gag ata cat ctg gga ccg gca gac agc 2425 Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser 195 200 205	5

											att Ile 220					2473
											acc Thr					2521
											gtg Val					2569
											gtg Val					2617
											cca Pro					2665
											gtc Val 300					2713
											tgc Cys					2761
ctt Leu	tac Tyr	ttg Leu	gtc Val	acg Thr 325	aag Lys	cat His	gcc Ala	gat Asp	gtc Val 330	att Ile	ccg Pro	gtg Val	cgc Arg	cgg Arg 335	cgg Arg	2809
											ccc Pro					2857
											tcg Ser					2905
											gtt Val 380					2953
											atg Met					3001
											cag Gln					3049
											agc Ser					3097

gcg Ala									3145
gcc Ala 450									3193
gac Asp									3241
atc Ile									3289
gly aaa									3337
tcg Ser									3385
gct Ala 530								gga Gly	3433
gtc Val									3481
gga Gly									3529
gj aaa									3577
ctc Leu			Gly	Leu	Leu				3625
cgg Arg 610									3673
gta Val									3721
gtg Val									3769

ctg Leu	gac Asp	ccg Pro	acc Thr 660	ttc Phe	acc Thr	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	gcg Ala	3817
gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	ggc 680	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	ggc	3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	ggc	atg Met	ttc Phe	gat Asp	3913
tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cys	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009
cca Pro	ej aaa	ttg Leu	ccc Pro 740	gtc Val	tgc Cys	cag Gln	gac Asp	cat His 745	ctg Leu	gag Glu	ttc Phe	tgg Trp	gag Glu 750	ggc	gtc Val	4057
Phe	Thr	ggc Gly 755	Leu	Thr	His	Ile	Asp 760	Ala	His	Phe	Leu	Ser 765	Gln	Thr	Lys	4105
Gln	Ala 770	gga Gly	Asp	Asn	Phe	Pro 775	Tyr	Leu	Val	Ala	Tyr 780	Gln	Ala	Thr	Val	4153
tgc Cys 785	'gcc Ala	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	aag Lys 800	4201
tgt Cys	ctc Leu	ata Ile	cgg Arg	cta Leu 805	aag Lys	cct Pro	acg Thr	ctg Leu	cac His 810	gly aaa	cca Pro	acg Thr	ccc Pro	ctg Leu 815	ctg Leu	4249
tat Tyr	agg Arg	ctg Leu	gga Gly 820	gcc Ala	gtt Val	caa Gln	aac Asn	gag Glu 825	gtt Val	act Thr	acc Thr	aca Thr	cac His 830	ccc Pro	ata Ile	4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc Gly 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gct Ala	gcg Ala	tat Tyr	4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441

•		agg Arg															4489
		gag Glu															4537
		cag Gln															4585
		gcc Ala 930															4633
	tgg Trp 945	cgg Arg	acc Thr	ctc Leu	gaa Glu	gcc Ala 950	ttc Phe	tgg Trp	gcg Ala	aag Lys	cat His 955	atg Met	tgg Trp	aat Asn	ttc Phe	atc Ile 960	4681
		gly aaa															4729
		ata Ile															4777
	acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	ej aaa	gga Gly 1009	\mathtt{Trp}	gtg Val	gcc Ala	4825
		caa Gln 1010	Leu					Ala					Val				4873
		gct Ala					Gly					Gly					4921
		att Ile		Ala	Gly	Tyr	Gly	Ala	Gly	Val	Ala		Ala	Leu		Ala	4969
		aag Lys			Ser					Ser					Val		5017
		ctc Leu		Ala					Gly					Gly			5065
		gca Ala 1090	Ala					His					Glu				5113

Gln 110	Trp 5	atg Met	Asn	Arg	Leu 111	Ile D	Ala	Phe	Ala	Ser 111	Arg 5	Gly	Asn	His	Val 1120	5161
tcc Ser	ccc Pro	acg Thr	cac His	tat Tyr 112	Val	cct Pro	gag Glu	agc Ser	gac Asp 113	Ala	gca Ala	gca Ala	cgt Arg	gtc Val 113!	Thr	5209
cag Gln	atc Ile	ctc Leu	tct Ser 1140	Ser	ctt Leu	acc Thr	atc Ile	act Thr 114	Gln	ctg Leu	ctg Leu	aag Lys	agg Arg 115	Leu	cac His	5257
cag Gln	tgg Trp	atc Ile 1155	Asn	gag Glu	gac Asp	tgc Cys	tcc Ser 1160	Thr	cca Pro	tgc Cys	tcc Ser	ggc Gly 116	Ser	tgg Trp	cta Leu	5305
aga Arg	gat Asp 1170	gtt Val O	tgg Trp	gat Asp	tgg Trp	ata Ile 1175	Cys	acg Thr	gtg Val	ttg. Leu	act Thr 118	Asp	ttc Phe	aag Lys	gcc Ala	5353
tgg Trp 1189	Leu	cag Gln	tcc Ser	aag Lys	ctc Leu 1190	Leu	ccg Pro	cga Arg	ttg Leu	ccg Pro 1199	Gly	gtc Val	ccc Pro	ttc Phe	ttc Phe 1200	5401
tca Ser	tgt Cys	caa Gln	cgt Arg	999 Gly 120	Tyr	aag Lys	gga Gly	gtc Val	tgg Trp 1210	Arg	ggc	gac Asp	Gly	atc Ile 1215	Met	5449
caa Gln	acc Thr	acc Thr	tgc Cys 1220	Pro	tgt · Cys	gga Gly	gca Ala	cag Gln 1225	Ile	acc Thr	gga Gly	cat His	gtg Val 1230	Lys	aac Asn	5497
tgt Cys	tcc Ser	atg Met 1235	Arg	atc Ile	gtg Val	gjå aaa	cct Pro 1240	Arg	acc Thr	tgt Cys	agt Ser	aac Asn 1245	Thr	tgg Trp	cat His	5545
gga Gly	aca Thr 1250	ttc Phe	ccc Pro	att Ile	aac Asn	gcg Ala 1255	Tyr	acc Thr	acg Thr	ggc Gly	ccc Pro 1260	Cys	acg Thr	ccc Pro	tcc Ser	5593
ccg Pro 1265	Ala	cca Pro	aat Asn	tat Tyr	tct Ser 1270	Arg	gcg Ala	ctg Leu	tgg Trp	cgg Arg 1275	Val	gct Ala	gct Ala	gag Glu	gag Glu 1280	5641
tac Tyr	gtg Val	gag Glu	Val	acg Thr 1285	Arg	gtg Val	eja aaa	gat Asp	ttc Phe 1290	His	tac Tyr	gtg Val	acg Thr	ggc Gly 1295	Met	5689
acc Thr	act Thr	gac Asp :	aac Asn 1300	Val	i,ys Lys	tgc Cys	ccg Pro	tgt Cys 1305	Gln	gtt Val	ccg Pro	gcc Ala	ccc Pro 1310	Glu	ttc Phe	5737
ttc Phe	Thr	gaa Glu 1315	Val	gat Asp	gjy aaa :	Val	cgg Arg 1320	Leu	cac His	agg Arg	tac Tyr	gct Ala 1325	Pro	gcg Ala	tgc Cys	5785

aaa ccc ctc cta Lys Pro Leu Leu 1330				
tac ccg gtt ggg Tyr Pro Val Gly 1345	tca cag ctc o Ser Gln Leu I 1350	Pro Cys Glu Pr	c gaa ctg gac o Glu Leu Asp 555	gta gca 5881 Val Ala 1360
gtg ctc act tcc Val Leu Thr Ser				
gct aag cgt agg Ala Lys Arg Arg 138	Leu Ala Arg (Ser Ser
tca gct agc cag Ser Ala Ser Glm 1395	Leu Ser Ala l			
cgt cat gac tcc Arg His Asp Ser 1410				
cgg cag gag atg Arg Gln Glu Met 1425		Ile Thr Arg Va		
gta gta att ttg Val Val Ile Leu				
agg gaa gta tcc Arg Glu Val Ser 146	Val Pro Ala (Lys Phe
cct cga gcg atg Pro Arg Ala Met 1475	Pro Ile Trp			
tta gag tcc tgg Leu Glu Ser Trp 1490		Asp Tyr Val Pr	o Pro Val Val	
			1200	
tgt cca ttg ccg Cys Pro Leu Pro 1505		gcc cct ccg at Ala Pro Pro Il	a cca cct cca	
Cys Pro Leu Pro	Pro Ala Lys in 1510 gtc ctg tca s	gcc cct ccg at Ala Pro Pro Il 15 gaa tct acc gt	ea cca cct cca le Pro Pro Pro 515 eg tct tct gcc	Arg Arg 1520 ttg gcg 6409

agc Ser	Gly	acg Thr 155	Ala	acg Thr	gcc Ala	tct Ser	cct Pro 1560	Asp	cag Gln	ccc Pro	tcc Ser	gac Asp 156	Asp	Gly	gac Asp	6505
gcg Ala	gga Gly 1570	Ser	gac Asp	gtt Val	gag Glu	tcg Ser 1579	Tyr	tcc Ser	tcc Ser	atg Met	ccc Pro 1580	Pro	ctt Leu	gag Glu	ej aaa	6553
gag Glu 158	ccg Pro 5	gl ^à aaa	gat Asp	ccc Pro	gat Asp 1590	Leu	agc Ser	gac Asp	eja aaa	tct Ser 1599	Trp	tct Ser	acc Thr	gta Val	agc Ser 1600	6601
gag Glu	gag Glu	gct Ala	agt Ser	gag Glu 160	Asp	gtc Val	gtc Val	tgc Cys	tgc Cys 1610	Ser	atg Met	tcc Ser	tac Tyr	aca Thr 1615	Trp	6649
aca Thr	ggc	gcc Ala	ctg Leu 1620	Ile	acg Thr	cca Pro	tgc Cys	gct Ala 1625	Ala	gag Glu	gaa Glu	acc Thr	aag Lys 1630	Leu	ccc Pro	6697
	aat Asn		Leu					Leu					Leu			6745
gct Ala	aca Thr 1650	Thr	tct Ser	cgc Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	6793
gac Asp 1665	aga Arg	ctg Leu	cag Gln	gtc Val	ctg Leu 1670	qzA	gac Asp	cac His	tac Tyr	cgg Arg 1675	Asp	gtg Val	ctc Leu	aag Lys	gag Glu 1680	6841
atg Met	aag Lys	gcg Ala	aag Lys	gcg Ala 1685	Ser	aca Thr	gtt Val	aag Lys	gct Ala 1690	Lys	ctt Leu	cta Leu	tcc Ser	gtg Val 1695	Glu	6889
gaa Glu	gcc Ala	tgt Cys	aag Lys 1700	Leu	acg Thr	ccc Pro	cca Pro	cat His 1705	Ser	gcc Ala	aga Arg	tct Ser	aaa Lys 1710	Phe	ggc Gly	6937
tat Tyr	gly ggg	gca Ala 1715	Lys	gac Asp	gtc Val	cgg Arg	aac Asn 1720	Leu	tcc Ser	agc Ser	aag Lys	gcc Ala 1725	Val	aac Asn	cac His	6985
atc Ile	cgc Arg 1730	Ser	gtg Val	tgg Trp	aag Lys	gac Asp 1735	Leu	ctg Leu	gaa Glu	gac Asp	act Thr 1740	Glu	aca Thr	cca Pro	att Ile	7033
gac Asp 1745	Thr	acc Thr	atc Ile	atg Met	gca Ala 1750	Lys	aat Asn	gag Glu	gtt Val	ttc Phe 1755	Cys	gtc Val	caa Gln	cca Pro	gag Glu 1760	7081
aag Lys	gjå aaa	ggʻc Gly	cgc Arg	aag Lys 1765	Pro	gct Ala	cgc Arg	ctt Leu	atc Ile 1770	Val	ttc Phe	cca Pro	gat Asp	ttg Leu 1775	Gly	7129

68 / 93

gtt (Val 2			Cys	Glu				Leu	Tyr				Ser	Thr		7177
			1780					1785					1790			
Pro (Val					Tyr					Ser			7225
cag (Gln :		Val					Asn					Lys				7273
atg g Met (1825	Gly	ttc Phe	gca Ala	tat Tyr	gac Asp 1830	Thr	cgc Arg	tgt Cys	ttt Phe	gac Asp 1835	Ser	acg Thr	gtc Val	act Thr	gag Glu 1840	7321
aat (Asn)					Glu					Gln					Ala	7369
ccc (_	-	_	Gln	_			-	Leu					Tyr		7417
Gly (Leu					Gly					Tyr			7465
tgc Cys		Ala					Thr					Asn				7513
tgt Cys 1905	Tyr					Ala					Ala					7561
tgc : Cys '					Cys					Val					Ser	7609
gcg (Glu					Leu					Glu		7657
atg Met			Tyr					Gly					Pro			7705
gac Asp		Glu					Cys					Ser				7753
gat Asp	Ala					Val					Arg					7801

ì

				gct Ala 2005	Ala					Arg					Asn	7849
				aac Asn)					Ala					Āla		7897
			Met	act Thr				Ser					Gln			7945
		Lys		cta Leu			Gln					Cys				7993
	Pro		_	cta Leu		Gln				_	Leu				_	8041
				cat His 2089	Ser					Glu					Ala	8089
				aaa Lys)					Pro					Arg		8137
cgg Arg	gcc Ala	aga Arg 2115	Ser	gtc Val	cgc Arg	gct Ala	agg Arg 2120	Leu	ctg Leu	tcc Ser	cag Gln	999 Gly 2129	Gly	agg Arg	gct Ala	8185
gcc Ala	act Thr 2130	Cys	ggc Gly	aag Lys	tac Tyr	ctc Leu 2135	Phe	aac Asn	tgg Trp	gca Ala	gta Val 2140	Arg	acc Thr	aag Lys	ctc Leu	8233
aaa Lys 2145	Leu	act Thr	cca Pro	atc Ile	ccg Pro 2150	Ala	gcg Ala	tcc Ser	cag Gln	ttg Leu 215	Asp	tta Leu	tcc Ser	agc Ser	tgg Trp 2160	8281
		Ala	Gly	tac Tyr 2169	Ser	Gly	Gly	Asp	Ile		His				Arg.	8329
gcc Ala	cga Arg	ccc Pro	cgc Arg 218(tgg Trp)	ttc Phe	atg Met	tgg Trp	tgc Cys 2185	Leu	ctc Leu	cta Leu	ctt Leu	tct Ser 2190	Val	gly aaa	8377
			Tyr	cta Leu				Arg		acgg	ggag	gct a	aaca	actco	a	8427
tttt ccat	tttt ctta	tt t	tttc	ctcct	t tt	tago	tect	ctt aaa	tttt	tcc	tttt	cttt	cc t	ttgg	ttttt tggct agaga	8547

<	211> 212> 213>	DNA						
	400> ccago						•	6
< <	210> 211> 212> 213>	63 DNA						
g	400> aatto ac		tggcgcgccc	agatgttaac	cagatccatg	gcacactcta	gagtactgtc	60 63
< <	210> 211> 212> 213>	33 DNA						
	400> ggaat		aacagaccac	aacggtttcc	ctc			33
< <	210> 211> 212> 213>	30 DNA						
	400> gcgta		tggtattatc	gtgttttca				30
< <	210> 211> 212> 213>	45 DNA						
	400> catat		tctaatacga	ctcactatag	gccagccccc	gattg		45
<: <:	210> 211> 212> 213>	45 DNA						
	400> gegee		ttggtttttc	tttgaggttt	aggattcgtg	ctcat		45
<: <:	210> 211> 212> 213>	36 DNA						
	400> aaggg		tgattgaaca	agatggattg	cacgca			36
	210> 211>							

<212> DNA <213> HCV	
<400> 15 gcatatgtta actcagaaga actcgtcaag aaggcgata	39
<210> 16	
<211> 45	
<212> DNA	
<213> HCV	
<400> 16	
gcatatgaat tctaatacga ctcactatag gccagccccc gattg	45
<210> 17	
<211> 30	
<212> DNA	
<213> HCV	
<400> 17	
acgcagaaag cgtctagcca tggcgttagt	30
<210> 18	
<211> 30	
<212> DNA	
<213> HCV	
<400> 18	
tcccggggca ctcgcaagca ccctatcagg	30
<210> 19	
<211> 26	
<212> DNA	
<213> HCV	
<220>	
<223> Label with FAM: fluorescence reporter dye	
<223> Label with TAMRA: Quencher dye	
<400> 19	
tggtctgcgg aacgggtgag tacacc	26
<210> 20	
<211> 45	
<212> DNA	
<213> HCV	
<400> 20	
gtggacgaat tctaatacga ctcactataa ccagcccccg attgg	45
<210> 21	
<211> 27	
<212> DNA	
<213> HCV	
<400> 21	
ggaacgcccg tcgtggccag ccacgat	27

```
<210> 22
<211> 23
<212> DNA
<213> HCV
<400> 22
gtcgtcttct ctgacatgga gac
                                                                  23
<210> 23
<211> 27
<212> DNA
<213> HCV
<400> 23
gagttgctca gtggattgat gggcagc
                                                                  27
<210> 24
<211> 8638
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 24
accagecece gattggggge gacactecae catagateae teccetgtga ggaactaetg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag eeteeaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teettettg gateaacccg eteaatgeet ggagatttgg gegtgeecc 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
eggeegettg ggtggagagg etattegget atgaetggge acaacagaca ateggetget 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacca ccaagegaaa categcateg agegageaeg tacteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg eccgaeggeg aggatetegt egtgaeecat ggegatgeet 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agcgcatcgc cttctatcgc cttcttgacg agttcttctg agttcgcgcc cagatgttaa 1200
cagaccacaa cggtttccct ctagogggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtett teceeteteg ecaaaggaat geaaggtetg ttgaatgteg tgaaggaage 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaaccccca cetggegaca ggtgeetetg eggecaaaag ceaegtgtat aagatacace 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggetetee teaagegtat teaacaaggg getgaaggat geecagaagg taccecattg 1680
tatgggatet gatetgggge eteggtgeae atgetttaea tgtgtttagt egaggttaaa 1740
aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac egg gag atg gea gea teg tge gga gge geg git tte gta ggt 1849
```

	et A	sp A	rg G		et Ai 5	la A	la Se	er Cy		ly G] 10	ly Al	la Va	al Pl		al Gly 15	•
	ata Ile															1897
	ata Ile															1945
caa Gln	gtg Val 50	tgg Trp	atc Ile	ccc Pro	ccc Pro	ctc Leu 55	aac Asn	gtt Val	cgg Arg	ejå aaa	60 Gly ggc	cgc Arg	gat Asp	gcc Ala	gtc Val	1993
	ctc Leu															2041
	atc Ile															2089
	acc Thr															2137
	atg Met															2185
	atg Met 130															2233
acc Thr 145	cca Pro	ctg Leu	cgg Arg	gac Asp	tgg Trp 150	gcc Ala	cac His	gcg Ala	ggc	cta Leu 155	cga Arg	gac Asp	ctt Leu	gcg Ala	gtg Val 160	2281
	gtt Val															2329
	Gly 999															2377
	tcc Ser															2425
	gaa Glu 210															2473
	cag Gln															2521

				gtc Val										2569
				gcg Ala										2617
				tca Ser										2665
				aat Asn										2713
				tcc Ser 310										2761
				agg Arg										2809
				agc Ser										2857
_		_		ggt Gly		_		_		_		_		2905
				gcc Ala									gtg Val	2953
_	-		_	gag Glu 390		_	_			_		_	-	3001
				tcc Ser	Pro	Pro	Ala	Val	Pro		Thr			3049
				cct Pro										3097
				caa Gln										3145
				ggt Gly										3193

									acg Thr			3241
									ggt Gly			3289
									cac His 510			3337
	Ser								caa Gln			3385
									cct Pro			3433
									ctg Leu			3481
									gag Glu			3529
_		 			_		_	_	aaa Lys 590	_	_	3577
									gta Val			3625
									gac Asp			3673
									gat Asp			3721
									gac Asp			3769
									caa Gln 670			3817
		Ser							agg Arg			3865

att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	ggc Gly	atg Met	ttc Phe	gat Asp	3913
tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cys	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009
cca Pro	gl ^a aaa	ttg Leu	ccc Pro 740	gtc Val	tgc Cys	cag Gln	gac Asp	cat His 745	ctg Leu	gag Glu	ttc Phe	tgg Trp	gag Glu 750	agc Ser	gtc Val	4057
		ggc Gly 755														4105
cag Gln	gca Ala 770	gga Gly	gac Asp	aac Asn	ttc Phe	ccc Pro 775	tac Tyr	ctg Leu	gta Val	gca Ala	tac Tyr 780	cag Gln	gct Ala	acg Thr	gtg Val	4153
tgc Cys 785	gcc Ala	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	aag Lys 800	4201
tgt Cys	ctc Leu	ata Ile	cgg Arg	cta Leu 805	aag Lys	cct Pro	acg Thr	ctg Leu	cac His 810	gly aaa	cca Pro	acg Thr	ccc Pro	ctg Leu 815	ctg Leu	4249
tat Tyr	agg Arg	ctg Leu	gga Gly 820	gcc Ala	gtt Val	caa Gln	aac Asn	gag Glu 825	gtt Val	act Thr	acc Thr	aca Thr	cac His 830	ccc Pro	ata Ile	4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	ġac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc Gly 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gcc Ala	gcg Ala	tat Tyr	4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc Gly	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	ttc Phe	4489
gat Asp	gag Glu	atg Met	gaa Glu 900	gag Glu	tgc Cys	gcc Ala	tca Ser	cac His 905	ctc Leu	cct Pro	tac Tyr	atc Ile	gaa Glu 910	cag Gln	gga Gly	4537

Met	cag Gln	ctc Leu 915	gcc Ala	gaa Glu	caa Gln	ttc Phe	aaa Lys 920	cag Gln	aag Lys	gca Ala	atc Ile	999 Gly 925	ttg Leu	ctg Leu	caa Gln	4585
	gcc Ala 930															4633
tgg Trp 945	cgg Arg	acc Thr	ctc Leu	gaa Glu	gcc Ala 950	ttc Phe	tgg Trp	gcg Ala	aag Lys	cat His 955	atg Met	tgg Trp	aat Asn	ttc Phe	atc Ile 960	4681
	gly aġa															4729
gcg Ala	ata Ile	gca Ala	tca Ser 980	ctg Leu	atg Met	gca Ala	ttc Phe	aca Thr 985	gcc Ala	tct Ser	atc Ile	acc Thr	agc Ser 990	ccg Pro	ctc Leu	4777
acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	gjå aaa	gga Gly 1005	\mathtt{Trp}	gtg Val	gcc Ala	4825
	caa Gln 1010	Leu					Ala					Val				4873
atc	gct	gga	aca	act	att	aaa	200	2+2	~~~							4001
Ile 1025	Ala	Gly	Ala	Ala	Val 1030	Gly	Ser	Ile	Gly	Leu 1035	Gly	aag Lys	Val	Leu	gtg Val 1040	4921
1025	Ala	Gly ttg	Ala gca	Ala ggt	Val 1030 tat Tyr	Gly gga	Ser gca	Ile 999	Gly gtg	Leu 1035 gca Ala	Gly	Lys	Val	Leu gtg	Val 1040 gcc Ala	4921
gat Asp	Ala att	Gly ttg Leu gtc	Ala gca Ala atg	ggt Gly 1045 agc Ser	Val 1030 tat Tyr	gga Gly gag	Ser gca Ala atg	Gly ggg ccc	gtg Val 1050 tcc Ser	Leu 1035 gca Ala	Gly ggc Gly	Lys gcg Ala gac	Val ctc Leu ctg	gtg Val 1055 gtt Val	Val 1040 gcc Ala	
gat Asp ttt Phe	att Ile	ttg Leu gtc Val	gca Ala atg Met 1060 gct Ala	ggt Gly 1045 agc Ser	Val 1030 tat Tyr ggc Gly	gga Gly gag Glu tcc	gca Ala atg Met	ggg Gly ccc Pro 1065 ggc Gly	gtg Val 1050 tcc Ser	gca Ala acc Thr	Gly ggc Gly gag Glu	Lys gcg Ala gac Asp	ctc Leu ctg Leu 1070	gtg Val 1055 gtt Val	Val 1040 gcc Ala aac Asn	4969
gat Asp ttt Phe cta Leu	Ala att Ile aag Lys	ttg Leu gtc Val cct Pro 1075	gca Ala atg Met 1060 gct Ala	ggt Gly 1045 agc Ser atc Ile	tat Tyr ggc Gly ctc Leu	gga Gly gag Glu tcc Ser	gca Ala atg Met cct Pro 1080	ggg ggc ggc ggc gtg	gtg Val 1050 tcc Ser gcc Ala	gca Ala acc Thr cta Leu	ggc Gly gag Glu gtc Val	gcg Ala gac Asp gtc Val 1085 gag Glu	ctc Leu ctg Leu 1070 ggg Gly	gtg Val 1055 gtt Val gtc Val	Val 1040 gcc Ala aac Asn gtg Val	4969 5017
gat Asp ttt Phe cta Leu tgc Cys	att Ile aag Lys ctc Leu gca Ala 1090 tgg	ttg Leu gtc Val cct Pro 1075 gcg Ala	gca Ala atg Met 1060 gct Ala ata Ile	ggt Gly 1045 agc Ser atc Ile ctg Leu	Val 1030 tat Tyr ggc Gly ctc Leu cgt Arg	gga Gly gag Glu tcc Ser cgg Arg 1095	gca Ala atg Met cct Pro 1080 cac His	ggg Gly ccc Pro 1065 ggc Gly gtg Val	gtg Val 1050 tcc Ser gcc Ala ggc Gly	gca Ala acc Thr cta Leu cca Pro	ggc Gly gag Glu gtc Val ggg Gly 1100 cgg	gcg Ala gac Asp gtc Val 1085 gag Glu	ctc Leu ctg Leu 1070 ggg Gly i	gtg Val 1055 gtt Val gtc Val gct Ala	Val 1040 gcc Ala aac Asn gtg Val gtg	4969 5017 5065

		g ctg ctg aag agg ctt c n Leu Leu Lys Arg Leu H 1150	
		a tgc tcc ggc tcg tgg c o Cys Ser Gly Ser Trp L 1165	
		g ttg act gat ttc aag a l Leu Thr Asp Phe Lys T 1180	
		g ccg gga gtc ccc ttc t u Pro Gly Val Pro Phe P 1195 1	
	Tyr Lys Gly Val Tr	g cgg ggc gac ggc atc a p Arg Gly Asp Gly Ile M 10 1215	
		c acc gga cat gtg aaa a e Thr Gly His Val Lys A 1230	
		c tgt agt aac acg tgg c r Cys Ser Asn Thr Trp H 1245	
		g ggc ccc tgc acg ccc t r Gly Pro Cys Thr Pro S 1260	
		g cgg gtg gct gct gag g p Arg Val Ala Ala Glu G 1275	
	Arg Val Gly Asp Ph	c cac tac gtg acg ggc a e His Tyr Val Thr Gly M 90 1295	
		g gtt ceg gee eee gaa t n Val Pro Ala Pro Glu P 1310	
ttc aca gaa gtg gat Phe Thr Glu Val Asy 1315	ggg gtg cgg ttg ca Gly Val Arg Leu Hi 1320	c agg tac gct cca gcg t s Arg Tyr Ala Pro Ala C 1325	gc 5785 Cys
		c ctg gtc ggg ctc aat c e Leu Val Gly Leu Asn G 1340	

gtg Val	ctc Leu	act Thr	tcc Ser	atg Met 136	Leu	acc Thr	gac Asp	ccc Pro	tcc Ser 137	His	att Ile	acg Thr	gcg Ala	gag Glu 137	Thr	5929
gct Ala	aag Lys	cgt Arg	agg Arg 138	Leu	gcc Ala	agg Arg	gga Gly	tct Ser 138	Pro	ccc Pro	tcc Ser	ttg Leu	gcc Ala 139	Ser	tca Ser	5977
tca Ser	gct Ala	agc Ser 1399	Gln	ctg Leu	tct Ser	gcg Ala	cct Pro 140	Ser	ttg Leu	aag Lys	gca Ala	aca Thr 140	Cys	act Thr	acc Thr	6025
cgt Arg	cat His 141	Asp	tcc Ser	ccg Pro	gac Asp	gct Ala 141!	Asp	ctc Leu	atc Ile	gag Glu	gcc Ala 142	Asn	ctc Leu	ctg Leu	tgg Trp	6073
cgg Arg 142	cag Gln 5	gag Glu	atg Met	gly	999 Gly 1430	Asn	atc Ile	acc Thr	cgc Arg	gtg Val 143	Glu	tca Ser	gaa Glu	aat Asn	aag Lys 1440	6121
gta Val	gta Val	att Ile	Leu	gac Asp 1445	Ser	ttc Phe	gag Glu	ccg Pro	ctc Leu 145(Gln	gcg Ala	gag Glu	gag Glu	gat Asp 1459	Glu	6169
agg Arg	gaa Glu	gta Val	tcc Ser 1460	Val	ccg Pro	gcg Ala	gag Glu	atc Ile 1465	Leu	cgg Arg	agg Arg	tcc Ser	agg Arg 1470	Lys	ttc Phe	6217
cct Pro	cga Arg	gcg Ala 1475	Met	ccc Pro	ata Ile	tgg Trp	gca Ala 1480	Arg	ccg Pro	gat Asp	tac Tyr	aac Asn 1485	Pro	cca Pro	ctg Leu	6265
tta Leu	gag Glu 1490	Ser	tgg Trp	aag Lys	gac Asp	ccg Pro 1495	Asp	tac Tyr	gtc Val	cct Pro	cca Pro 1500	Val	gta Val	cac His	gl ^à aaa	6313
tgt Cys 1505	cca Pro	ttg Leu	ccg Pro	cct Pro	gcc Ala 1510	Lys	gcc Ala	cct Pro	ccg Pro	ata Ile 1515	Pro	cct Pro	cca Pro	cgg Arg	agg Arg 1520	6361
aag Lys	agg Arg	acg Thr	Val	gtc Val 1525	Leu	tca Ser	gaa Glu	tct Ser	acc Thr 1530	Val	tct Ser	tct Ser	gcc Ala	ttg Leu 1535	Ala	6409
gag Glu	ctc Leu	gcc Ala	aca Thr 1540	Lys	acc Thr	ttc Phe	ggc Gly	agc Ser 1545	Ser	gaa Glu	tcg Ser	tcg Ser	gcc Ala 1550	Val	gac Asp	6457
agc Ser	ggc Gly	acg Thr 1555	Ala	acg Thr	gcc Ala	tct Ser	cct Pro 1560	Asp	cag Gln	ccc Pro	tcc Ser	gac Asp 1565	Asp	ggc Gly	gac Asp	6505
gcg Ala	gga Gly 1570	Ser	gac Asp	gtt Val	Glu	tcg Ser 1575	Tyr	tcc Ser	tcc Ser	atg Met	ccc Pro 1580	Pro	ctt Leu	gag Glu	ejå aaa	6553

			tgg tct acc gta agc 6601 Trp Ser Thr Val Ser 1600
			atg tcc tac aca tgg 6649 Met Ser Tyr Thr Trp 1615
	Ile Thr Pro		gaa acc aag ctg ccc 6697 Glu Thr Lys Leu Pro 1630
	Ser Asn Ser	_	cac aac ttg gtc tat 6745 His Asn Leu Val Tyr 1645
			aag aag gtc acc ttt 6793 Lys Lys Val Thr Phe 1660
		•	gac gtg ctc aag gag 6841 Asp Val Leu Lys Glu 5 1680
			ctt cta tcc gtg gag 6889 Leu Leu Ser Val Glu 1695
	Leu Thr Pro		aga tct aaa ttt ggc 6937 Arg Ser Lys Phe Gly 1710
	Asp Val Arg	_	aag gcc gtt aac cac 6985 Lys Ala Val Asn His 1725
			act gag aca cca att 7033 Thr Glu Thr Pro Ile 1740
Asp Thr Thr Ile	Met Ala Lys		tgc gtc caa cca gag 7081 Cys Val Gln Pro Glu 5 1760
			ttc cca gat ttg ggg 7129 Phe Pro Asp Leu Gly 1775
	Glu Lys Met		gtg gtc tcc acc ctc 7177 Val Val Ser Thr Leu 1790

		Val	gag Glu				Asn					Lys				7273
	Gly		gca Ala			Thr					Ser					7321
			cgt Arg		Glu					Gln					Ala	7369
			aga Arg 1860	Gln					Leu					Tyr		7417
ejå aaa	Gly	ccc Pro 1875	ctg Leu 5	act Thr	aat Asn	tct Ser	aaa Lys 1880	Gly	cag Gln	aac Asn	tgc Cys	ggc Gly 1885	${\tt Tyr}$	cgc Arg	Arg Cgg	7465
		Ala	agc Ser				Thr					Asn				7513
	Tyr		aag Lys			Ala					Ala					7561
			ctc Leu		Cys										Ser	7609 _.
Cys	Thr ggg	Met acc		Val 1925 gaġ Glu	Cys gac	Gly gag	Asp gcg	Asp agc	Leu 1930 cta Leu	Val) cgg	Val gcc	Ile ttc	Cys	Glu 1935 gag Glu	Ser gct	7609 _.
Cys gcg Ala atg	Thr ggg Gly act	Met acc Thr	caa Gln 1940 tac Tyr	Val 1925 gaġ Glu) tct	Gac Asp	gag Glu ccc	Asp gcg Ala cct	agc Ser 1945 ggg Gly	Leu 1930 cta Leu gac	val cgg Arg	Val gcc Ala ccc	Ile ttc Phe	Cys acg Thr 1950 cca Pro	Glu 1935 gag Glu)	gct Ala	•
gcg Ala atg Met	Thr 999 Gly act Thr	Met acc Thr aga Arg 1955 gag Glu	caa Gln 1940 tac Tyr ttg Leu	yal 1925 gaġ Glu tct Ser ata Ile	gac Asp gcc Ala aca Thr	gag Glu ccc Pro	gcg Ala cct Pro 1960 tgc Cys	agc ser 1945 ggg Gly tcc ser	tcc Ser	cgg Arg ccg Pro	yal gcc Ala ccc Pro	ttc Phe aaa Lys 1965 tca Ser	acg Thr 1950 cca Pro	Glu 1935 gag Glu gaa Glu	gct Ala tac Tyr	7657
gcg Ala atg Met gac Asp	ggg Gly act Thr ttg Leu 1970 gca Ala	Met acc Thr aga Arg 1955 gag Glu	caa Gln 1940 tac Tyr ttg Leu	yal 1925 gaġ Glu tct Ser ata Ile	gac Asp gcc Ala aca Thr	gag Glu ccc Pro tca Ser 1975 gtg Val	gcg Ala cct Pro 1960 tgc Cys	agc Ser 1945 999 Gly tcc Ser	tcc ser	cgg Arg ccg Pro aat Asn	yal ccc Pro gtg Val 1980 cgt Arg	ttc Phe aaa Lys 1965 tca Ser	acg Thr 1950 cca Pro gtc Val	gag Glu gaa Glu gaa Glu gcg Ala	gct Ala tac Tyr cac His	7657 7705
gcg Ala atg Met gac Asp gat Asp 1985	ggg Gly act Thr ttg Leu 1970 gca Ala	Met acc Thr aga Arg 1955 gag Glu tct Ser gcg	caa Gln 1940 tac Tyr ttg Leu	yal 1925 gaġ Glu tct Ser ata Ile aaa Lys	gac Asp gcc Ala aca Thr agg Arg 1990 gcg Ala	gag Glu ccc Pro tca Ser 1975 gtg Val	gcg Ala cct Pro 1960 tgc Cys tac Tyr	agc Ser 1945 ggg Gly tcc Ser tat Tyr	tcc ser ctc Leu	cgg Arg ccg Pro aat Asn acc Thr 1995	yal ccc pro gtg Val 1980 cgt Arg	ttc Phe aaa Lys 1965 tca Ser gac Asp	acg Thr 1950 cca Pro Stc Val ccc Pro	gag Glu gaa Glu gcg Ala acc Thr	gct Ala tac Tyr cac His acc Thr 2000 aat Asn	7657 7705 7753

62, 66	
atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045	;
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993 Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2050 2055 2060	,
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 2080	•
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095)
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	,
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	;
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	3
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	L
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175)
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190	7
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427 Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	7
ggccaatagg ccatcetgtt titttecett titttitte titttitt titttitt 8487 titttittit titteett tittteet ettitite tittette tittgggget 8547 ccatcitage cctagicaeg getagetgtg aaaggicegt gageegetig aetgeagaga 8607 gigetgatae tggcctetet geagateaag t	7
<210> 25 <211> 8638 <212> DNA <213> HCV	
<220> <221> CDS <222> (1802)(8407)	

		> 25															
																ctactg	
																caggac	
																tgccag	
																acccc	
																gatagg	
																ctaaac	
																gttctc	
C	ggc	cgct	tg 9	ggtgg	gagag	g ct	tatto	cggct	ate	gact	gggc	acaa	acaga	aca a	atcgg	gctgct	480
																agaccg	
																tggcca	
																actggc	
																ccgaga	
aa	gta	atco	cat o	catgo	gctga	at go	caat	gegge	gge	ctgc	atac	gctt	tgato	ccg (gctad	cctgcc	780
																ccggtc	
ti	gto	cgat	ca g	ggato	gatct	g ga	acgaa	agago	ato	cagg	ggct	cgcg	gecag	gee 9	gaact	tgttcg	900
																atgcct	
g	ette	gccg	gaa 1	tatca	atggt	g ga	aaaa	tggcc	gct	tttt	ctgg	atto	catc	gac	tgtg	geegge	1020
tg	gggt	tgtg	gc s	ggac	gcta	at ca	agga	catag	r cgt	ttgg	ctac	ccgt	gata	att (gctga	aagagc	1080
tt	ggd	cggc	ga a	atggg	gctga	ac cg	gctto	cctcg	, tg	cttta	acgg	tato	geeg	gct	cccga	attcgc	1140
																tgttaa	
																tactgg [,]	
C	gaa	agco	gc t	tgga	aataa	ag go	ccggt	tgtgc	gt!	ttgt	ctat	atgt	tatt	tt	ccac	catatt	1320
																cattcc	
																ggaagc	
																gcagcg	
ga	acc	ccc	ca	cctg	gcgad	a g	gtgc	ctctg	g	gccaa	aaag	cca	egtgt	tat	aagai	tacacc	1560
	,		,~9 :	guave	aaccc		9 - 9 - 1	ے درجات					46696	-44	aaaqa	agtcaa	1020
tg																agtcaa ccattq	
t g at	ggc	ctct	cc t	caag	gcgta	at to	caaca	aaggg	g gct	tgaag	ggat	gcc	cagaa	agg	tacco	ccattg	1680
to at	ggc	ctct ggat	cc (caac gatct	gegta :gggg	at to go ct	caaca ceggt	aaggg tgcac	gct ato	tgaag getti	ggat taca	gccc	cagaa gttta	agg agt	tacco cgago	ccattg gttaaa	1680 1740
to at ta aa	ggc tgg acg	etet ggat gtet	cc (ct (caaq gatet gece	gegta egggg eeega	at to go ct aa co	caaca ceggt caegg	aaggg tgcac gggac	get atg	tgaag getti ggtti	ggat taca tcc	gccc tgtg tttg	eagaa gttta gaaaa	agg agt aac	tacco cgago acgat	ccattg gttaaa taatac	1680 1740 1800
to at ta aa	ggo itgg iacg atg	etet ggat gtet g ga	cc (ct (ag (caag gatet geeed gg ga	gegta gggg geega ag at	at to go et aa co g go	caaca ceggt caegg ca go	aaggg tgcac gggac ca tc	get atg gtg	tgaag getti ggtti ge gg	ggat taca ttcc ga gg	geed tgtg tttg	cagaa gttta gaaaa cg gt	agg agt aac at t	tacco cgago acgat tc gi	ccattg gttaaa taatac ta ggt	1680 1740 1800
to at ta aa	ggo itgg iacg atg	etet ggat gtet g ga	cc (ct (ag (caag gatet geeed gg ga	gegta gggg geega ag at	at to go of aa co eg go et Al	caaca ceggt caegg ca go	aaggg tgcac gggac ca tc	get atg gtg	tgaag getti ggtti ge gg ys G:	ggat taca ttcc ga gg	geed tgtg tttg	cagaa gttta gaaaa cg gt	agg agt aac at t	tacco cgago acgal tc gi he Va	ccattg gttaaa taatac ta ggt al Gly	1680 1740 1800
to at ta aa	ggo atgg acg atg Met	etet ggat gtet g ga	cc (ct (ag (caag gatet geeed gg ga	gegta gggg eeega ag at lu Ma	at to go of aa co eg go et Al	caaca ceggt caegg ca go	aaggg tgcac gggac ca tc	get atg gtg	tgaag getti ggtti ge gg ys G:	ggat taca tcc ga gg ly G]	geed tgtg tttg	cagaa gttta gaaaa cg gt	agg agt aac at t	tacco cgago acgal tc gi he Va	ccattg gttaaa taatac ta ggt	1680 1740 1800
at ta aa c	ggo itgg iacg atg Met	etet ggat gtet g ga t As	cc (ct (ag (ac co	caag gatet geeed gg ga	gegta egggg ecega ag at lu Me	at to go of aa co eg go et Al	caaca ceggt caegg ca go la Al	aaggg tgcac gggac ca tc la Se	get ato gto g to er Cy	tgaag getti ggtti ge gg ys G:	ggat taca ttcc ga gg ly G]	geed tgtg tttg ge ge ly Al	cagaa gttta gaaaa cg gt la Va	agg agt aac at t al P	tacco cgago acgat tc gt he Va	ccattg gttaaa taatac ta ggt al Gly 15	1680 1740 1800
tg at ta aa c	ggo atgg atg atg Met	etet ggat gtet g ga t As	ctc	ttg	gegta gggg eeega ag at lu Me	at to go et aa co et Al 5	caaca ceggt caegg ca go la Al	aaggg tgcac gggac ca tc la Se	g got g ato g g to er Cy cac	tgaag getti ggtti ge gg ys G: :	ggat taca ttcc ga gg ly G] 10	geed tgtg ttttg ge ge ly Al	eagaa gttta gaaaa eg gt la Va	agg agt ac at t al P	tacco cgago acgat tc gi he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
tg at ta aa c	ggo atgg atg atg Met	etet ggat gtet g ga t As	ctc	ttg	gegta egggg ecega ag at lu Me	at to go et aa co et Al 5	caaca ceggt caegg ca go la Al	aaggg tgcac gggac ca tc la Se	g got g ato g g to er Cy cac	tgaag getti ggtti ge gg ys G: :	ggat taca ttcc ga gg ly G] 10	geed tgtg ttttg ge ge ly Al	eagaa gttta gaaaa eg gt la Va	agg agt ac at t al P	tacco cgago acgat tc gi he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
tg at ta aa c	ggo atgg atg atg Met	etet ggat gtet g ga t As	ctc	ttg Leu	gegta gggg eeega ag at lu Me	at to go et aa co et Al 5	caaca ceggt caegg ca go la Al	aaggg tgcac gggac ca tc la Se	get e ate e gte eg te er Cy cac	tgaag getti ggtti ge gg ys G: :	ggat taca ttcc ga gg ly G] 10	geed tgtg ttttg ge ge ly Al	eagaa gttta gaaaa eg gt la Va	agg agt act t al P ctc Leu	tacco cgago acgat tc gi he Va gct	ccattg gttaaa taatac ta ggt al Gly 15	1680 1740 1800 1849
te at aa c	eggo atgo atgo Met 1	etet ggat gtet g ga t As	ctc (ctc) ctc ctc ctc ctc Leu	ttg Leu 20	gegta gggg ceega ag at lu Me 5 acc Thr	at to go of aa co eg go et Al 5 ttg Leu	caaca ceggt caegg ca go la Al tea Ser	aaggg ggac gggac ca tc la Se ccg Pro	get get get er Cy cac His 25	tgaaggetti ggtti ge gg ys G: tat Tyr	ggat taca ttcc ga gg ly Gl 10 aag Lys	gccc tgtc tttc gc gc ly Al ctg Leu	ttc Phe	agg agt act tt al P ctc Leu 30	tacco cgagg acgal tc gt he Va gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
at ta aa c	eggo atgo atgo Met 1 eg a	etet ggat gtet gtet gte ata ata	ctc Leu	tcaag gatet geeco gg ga rg G] ttg Leu 20	gegta geggg ceega ag at lu Me 5 acc Thr	at to go of aa co eg go et Al i ttg Leu caa	caaca teggt caegg ca go la Al tea Ser	aaggg tgcac gggac ca tc la Se ccg Pro	get atg atg get car Cy cac His 25	tgaaggetti ggtti ge gg ys G: tat Tyr	ggat taca ttcc ga gg ly Gl 10 aag Lys	gccc tgtc tttc gc gc ly Al ctg Leu	ragaa gttta gaaaa rg gt la Va ttc Phe	agg agt act tt al P ctc Leu 30	tacco cgagg acgat tc gt he Va : gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
at ta aa c	eggo atgo atgo Met 1 eg a	etet ggat gtet gtet gte ata ata	ctc Leu	tcaag gatet geeco gg ga rg G] ttg Leu 20	gegta gggg ceega ag at lu Me 5 acc Thr	at to go of aa co eg go et Al i ttg Leu caa	caaca teggt caegg ca go la Al tea Ser	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe	get atg atg get car Cy cac His 25	tgaaggetti ggtti ge gg ys G: tat Tyr	ggat taca ttcc ga gg ly Gl 10 aag Lys	gccc tgtc tttc gc gc ly Al ctg Leu	gagaa gttta gaaaa cg gt la Va ttc Phe gag Glu	agg agt act tt al P ctc Leu 30	tacco cgagg acgat tc gt he Va : gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
at ta aa c	eggo atgo atgo Met 1 eg a	etet ggat gtet gtet gte ata ata	ctc Leu	tcaag gatet geeco gg ga rg G] ttg Leu 20	gegta geggg ceega ag at lu Me 5 acc Thr	at to go of aa co eg go et Al i ttg Leu caa	caaca teggt caegg ca go la Al tea Ser	aaggg tgcac gggac ca tc la Se ccg Pro	get atg atg get car Cy cac His 25	tgaaggetti ggtti ge gg ys G: tat Tyr	ggat taca ttcc ga gg ly Gl 10 aag Lys	gccc tgtc tttc gc gc ly Al ctg Leu	ragaa gttta gaaaa rg gt la Va ttc Phe	agg agt act tt al P ctc Leu 30	tacco cgagg acgat tc gt he Va : gct Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg	1680 1740 1800 1849
te at ta aa c	atggaatgaatgaatgaatgaatgaatgaatgaatgaat	ctct ggat gtct g ga t As ata Ile	ccc (cct (cct (cct (cct (cct (cct (cct	ttg Leu tgg Trp	gegta egggg ecega ag at lu Ma acc Thr tta Leu	ttg caa caa caa Gln	tca ser ttat	eaggg cgcac gggac ca to la Se ccg Pro ttt Phe 40	g get g atg g tg gr Cy cac His 25 atc	tgaaggetti ggtti ggtti ge gg ys Gi tat Tyr acc Thr	ggat taca ttcc ga gg ly GJ 10 aag Lys agg Arg	gcc tgtg tttg gc gc ctg Leu gcc Ala	ttc Phe gag Glu 45	ctc Leu 30 gca	tacco cgagg acgat tc gt he Va gct Ala cac	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu	1680 1740 1800 1849 1897
to at ta as a c c the Lee ct	ggcatggaatgaatgaatgaatgaatgaatgaatgaatga	ctetet ggat gtet ggat ggat ata ata ata ata	ctc tgg Trp 35	tcaaggatet geeec gg ga rg Gl ttg Leu 20 tgg Trp	gegta gggg ceega ag at lu Me g acc Thr tta Leu	ttg caa caa Gln	tca ser tat tat	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40	g get g atg g tg g tg cac His 25 atc Ile	tgaaggettiggttiggttigget	ggat taca ttcc ga gg ly Gl lo aag Lys agg Arg	gcc tgtg tttg gc gc ctg Leu gcc Ala	ttc Phe gag Glu 45	agg agt act tt tal P ctc Leu 30 gca Ala	taccoccgaggacgatcgatcgatcgatcgatcgatcgatcgatc	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849
to at ta as a c c the Lee ct	ggcatggaatgaatgaatgaatgaatgaatgaatgaatga	ctctctctctctctctctctctctctctctctctctct	ctc tgg Trp 35	tcaaggatet geeec gg ga rg Gl ttg Leu 20 tgg Trp	gegta egggg ecega ag at lu Ma acc Thr tta Leu	ttg caa caa Gln	tca ser tat Tyr	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40	g get g atg g tg g tg cac His 25 atc Ile	tgaaggettiggttiggttigget	ggat taca ttcc ga gg ly Gl lo aag Lys agg Arg	gcc tgtg tttg gc gc Leu gcc Ala	ttc Phe gag Glu 45	agg agt act tt tal P ctc Leu 30 gca Ala	taccoccgaggacgatcgatcgatcgatcgatcgatcgatcgatc	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849 1897
to at ta as a c c the Lee ct	ggcatggaatgaatgaatgaatgaatgaatgaatgaatga	ctetet ggat gtet ggat ggat ata ata ata ata	ctc tgg Trp 35	tcaaggatet geeec gg ga rg Gl ttg Leu 20 tgg Trp	gegta gggg ceega ag at lu Me g acc Thr tta Leu	ttg caa caa Gln	tca ser tat tat	aaggg tgcac gggac ca tc la Se ccg Pro ttt Phe 40	g get g atg g tg g tg cac His 25 atc Ile	tgaaggettiggttiggttigget	ggat taca ttcc ga gg ly Gl lo aag Lys agg Arg	gcc tgtg tttg gc gc ctg Leu gcc Ala	ttc Phe gag Glu 45	agg agt act tt tal P ctc Leu 30 gca Ala	taccoccgaggacgatcgatcgatcgatcgatcgatcgatcgatc	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc	1680 1740 1800 1849 1897
te at ta as c	eggcateggaacgaacgaacgaacgaacgaacgaacgaacgaacga	ctctctggat gggt ggt t As ata alle gtg 7al	ctc tgg tgg Trp tgg	ttg Leu tgg Trp	gegta gggg ceega ag at lu Ma acc Thr tta Leu	ttg caa caa Gln ccc	tca ser tat Tyr	aaggg ggac gggac ca tc la Se ccg Pro ttt Phe 40 aac Asn	g get e atc gtt cac cac 25 atc Ile	tgaaggetti ggtti ggtti ge gg ys G tat Tyr acc Thr	ggat caca ctcc ga gg ly Gl lo aag Lys agg Arg	gcc tgtg ttttg gc gc Leu gcc Ala ggc Gly	ttc gag tttc gag ttc Phe gag Glu 45	ctc Leu 30 gca Ala	taccoccyagos acgattc got Ala cac His	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1680 1740 1800 1849 1897 1945
tg at ta aa c c t Le c c	eggoatgeatgeatgeatgeatgeatgeatgeatgeatgeatge	ctctctcggat gggtctgggtct ggat t As ata alle gtg 7al 50 ctc	ctc tgg Trp Trp ctc	ttg Leu 20 tgg Trp atc	gegta geggg ceega ag at lu Me acc Thr tta Leu ccc Pro	ttg caa Gln ccc pcg	tca ser tat Tyr ctc Leu 55	cac	g get g tog g tog cac His 25 atc Ile gtt Val	tgaaggetti ggtti gg tg ys G; tat Tyr acc Thr cgg Arg	ggat caca ctcc ga gg ly Gl 10 aag Lys agg Arg Gly cta	gcc tgtg tttg gc gd y Al ctg Leu gcc Ala ggc Gly 60	ttc phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccyange acgaint cagaint cagaint cagaint cagaint caccing acgaint acgaint caccing acgaint caccing acgaint caccing acgaint a	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1680 1740 1800 1849 1897
tg at ta aa	eggoateggaategaategaategaategaategaatega	ctctctcggat gggtctgggtct ggat t As ata alle gtg 7al 50 ctc	ctc tgg Trp Trp ctc	ttg Leu 20 tgg Trp atc	gegta gggg ceega ag at lu Ma acc Thr tta Leu	ttg caa Gln ccc gcg Ala	tca ser tat Tyr ctc Leu 55	cac	g get g tog g tog cac His 25 atc Ile gtt Val	tgaaggetti ggtti gg tg ys G; tat Tyr acc Thr cgg Arg	ggat caca ctcc ga gg ly G] l0 aag Lys agg Arg ggg Gly cta Leu	gcc tgtg tttg gc gd y Al ctg Leu gcc Ala ggc Gly 60	ttc phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccyange acgaint cagaint cagaint cagaint cagaint caccing acgaint acgaint caccing acgaint caccing acgaint caccing acgaint a	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr	1680 1740 1800 1849 1897 1945
tg at ta aa	eggoatgeatgeatgeatgeatgeatgeatgeatgeatgeatge	ctctctcggat gggtctgggtct ggat t As ata alle gtg 7al 50 ctc	ctc tgg Trp Trp ctc	ttg Leu 20 tgg Trp atc	gegta geggg ceega ag at lu Me acc Thr tta Leu ccc Pro	ttg caa Gln ccc pcg	tca ser tat Tyr ctc Leu 55	cac	g get g tog g tog cac His 25 atc Ile gtt Val	tgaaggetti ggtti gg tg ys G; tat Tyr acc Thr cgg Arg	ggat caca ctcc ga gg ly Gl 10 aag Lys agg Arg Gly cta	gcc tgtg tttg gc gd y Al ctg Leu gcc Ala ggc Gly 60	ttc phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	taccoccyange acgaint cagaint cagaint cagaint cagaint caccing acgaint acgaint caccing acgaint caccing acgaint caccing acgaint a	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1680 1740 1800 1849 1897 1945
to at take the case of the cas	eggoateggateggateggateggateggateggategga	eteteteggat ggat ggat ggat ggat ata ata alle sta 50 etc eu	ctc Leu tgg Trp ctc Leu tgg	ttg Trp atc acg	gegta gggg ceega ag at lu Ma acc Thr tta Leu ccc Pro	ttg caa Gln ccc pro	tca ser tat Tyr ctc Leu 55 atc	cac His	g gct gtg gtg gtg gtg gtg gtg gtg gtg gt	tgaaggetti ggtti ggtti gc gg ys G: tat Tyr acc Thr cgg Arg	ggat caca ctcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75	gcc tgtg ttttg gc gd y Al ctg Leu gcc Ala ggc Gly 60 atc	ttc pagg ttta gaaaa ggt la Va ttc Phe gag Glu 45 cgc Arg	agg agt agt tal post of the ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyaggacgattc ggthe Vala cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1680 1740 1800 1849 1897 1945
ct Les Cat Les	eggoatggaatgaatgaatgaatgaatgaatgaatgaatga	ctct ggat ggat ggat ggat at As at a at a at a lle gtgl 50 ctc	ctc Leu tgg Trp ctc Leu ttg	ttg Leu 20 tgg Trp atc acg Thr	gegta gggg ceega ag at lu Ma acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro Ala 70 ata	tca ggt tat Tyr ctc Leu 555 atc Ile ctc	aggggggaca to la Se cog Pro ttt Phe 40 aac Asn cac His	g get g atc g tc cac His 25 atc Ile gtt Val	tgaag getti ggtti gg tg ys G: tat Tyr acc Thr cgg Arg gag Glu	ggat caca ctcc ga gg ly G] lo aag Lys agg Arg Gly cta Leu 75	gcc tgtg gc gc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	agg agt agt tal post of the ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyaggacgathe Value GCt Ala Cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1680 1740 1800 1849 1897 1945
ct Les Cat Les	eggoatggaatgaatgaatgaatgaatgaatgaatgaatga	ctct ggat ggat ggat ggat at As at a at a at a lle gtgl 50 ctc	ctc Leu tgg Trp ctc Leu ttg	ttg Leu 20 tgg Trp atc acg Thr	gegta gggg ceega ag at lu Me acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro Ala 70 ata	tca ggt tat Tyr ctc Leu 555 atc Ile ctc	aggggggaca to la Se cog Pro ttt Phe 40 aac Asn cac His	g get g atc g tc cac His 25 atc Ile gtt Val	tgaag getti ggtti gc gg ys G: tat Tyr acc Thr cgg Arg gag Glu ctc Leu	ggat caca ctcc ga gg ly G] lo aag Lys agg Arg Gly cta Leu 75	gcc tgtg gc gc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	agg agt agt tal post of the ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyagos acgardate get Ala cac Ala atc Ile get Ala	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1680 1740 1800 1849 1897 1945
ct Les Cat Les	eggoatggaatgaatgaatgaatgaatgaatgaatgaatga	ctct ggat ggat ggat ggat at As at a at a at a lle gtgl 50 ctc	ctc Leu tgg Trp ctc Leu ttg	ttg Leu 20 tgg Trp atc acg Thr	gegta gggg ceega ag at lu Ma acc Thr tta Leu ccc Pro tgc Cys	ttg caa Gln ccc pro Ala 70 ata	tca ggt tat Tyr ctc Leu 555 atc Ile ctc	aggggggaca to la Se cog Pro ttt Phe 40 aac Asn cac His	g get g atc g tc cac His 25 atc Ile gtt Val	tgaag getti ggtti gg tg ys G: tat Tyr acc Thr cgg Arg gag Glu	ggat caca ctcc ga gg ly G] lo aag Lys agg Arg Gly cta Leu 75	gcc tgtg gc gc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	agg agt agt tal post of the ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyaggacgathe Value GCt Ala Cac His gcc Ala atc Ile	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1680 1740 1800 1849 1897 1945
to at Le Carrier Le Ca	eggoatggaatgaatgaatgaatgaatgaatgaatgaatga	ctct ggat ggat ggat ggat ggat ata ata ata a	ctc Leu tgg Trp ctc Leu ttgg Leu ttgg	ttg Leu 20 tgg Trp atc acg	gegta geggg ceega ag at lu Me acc Thr tta Leu ccc Pro tgc Cys	ttg cet Alico caa Gln ccc pro ata Ile	tca ggt tat Tyr ctc Leu ctc Leu	agggggggacga tola Second Pro ttt Phe 40 aac Asn cac His	g get con cac His 25 atc Ile gtt Val cca Pro	tgaag getti ggtti gg gg ys G: tat Tyr acc Thr cgg Glu ctc Leu 90	ggat caca ctcc ga gg ly Gl lo aag Lys agg Arg ggg Gly cta Leu 75 atg	gcc tgtg y Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc phe gag Glu 45 cgc Arg ttt phe	ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyagos acgardate get Ala cac His get Ala atc Ile get Ala 95	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945 1993 2041
ct Les Cat Les	eggotace and set of the set of th	eteteteggat ggat ggat ggat ggat ggat gga	ctc Leu ttgg Trp ctc Leu ttgg Leu aaa	ttg Leu 20 tgg Trp atc Leu ctc Leu gtg	gegta geggg ceega ag at lu Ma acc Thr tta Leu ccc Pro tgc Cys	ttg cet Alici ce	tca ggt tac ggt tac ggt tac ser tat Tyr ctc Leu 55 atc Ile ctc Leu ttc	agggggggacggaca to a se ccg Pro ttt Phe 40 aac Asn cac His ggt Gly	g get cac cac His 25 atc Ile gtt Val cca Pro	tgaaggetti ggtti ggtti ggtgys gc ggys tat Tyr acc Thr cgg Glu ctc Leu 90 gca	ggat caca ctcc ga gg ly Gl lo aag Lys agg Gly cta Leu 75 atg Met	gcc tgtg gc gc Ala ggc Ala ggc Ala ggc Ala ggg Yal	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu ctc	ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyago acgat tc gg he Va gct Ala cac His gcc Ala atc Ile gct Ala 95 cgt	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945
ct Les Cat Les	eggotace and set of the set of th	eteteteggat ggat ggat ggat ggat ggat gga	ctc Leu ttgg Trp ctc Leu ttgg Leu aaa	ttg Leu 20 tgg Trp atc Leu ctc Leu gtg	gegta geggg ceega ag at lu Me acc Thr tta Leu ccc Pro tgc Cys	ttg cet Alici ce	tca ggt tac ggt tac ggt tac ser tat Tyr ctc Leu 55 atc Ile ctc Leu ttc	agggggggacggaca to a se ccg Pro ttt Phe 40 aac Asn cac His ggt Gly	g get cac cac His 25 atc Ile gtt Val cca Pro	tgaaggetti ggtti ggtti ggtgys gc ggys tat Tyr acc Thr cgg Glu ctc Leu 90 gca	ggat caca ctcc ga gg ly Gl lo aag Lys agg Gly cta Leu 75 atg Met	gcc tgtg gc gc Ala ggc Ala ggc Ala ggc Ala ggg Yal	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu ctc	ctc Leu 30 gca Ala gat Asp acc Thr	taccoccyagos acgarded to go the Value at a cac Ala at a c	ccattg gttaaa taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1680 1740 1800 1849 1897 1945 1993 2041

					aag Lys											2185
					gca Ala											2233
					tgg Trp 150											2281
					gtc Val											2329
					gcg Ala											2377
					ej aaa											2425
					tgg Trp											2473
					cta Leu 230											2521
cgg Arg	gac Asp	agg Arg	aac Asn	cag Gln 245	gtc Val	gag Glu	ejà aaa	gag Glu	gtc Val 250	caa Gln	gtg Val	gtc Val	tcc Ser	acc Thr 255	gca Ala	2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	gcg Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	gtc Val	2617
				Gly	tca Ser	Lys	Thr	Leu		Gly		Lys	Gly			2665
acc Thr	caa Gln 290	atg Met	tac Tyr	acc Thr	aat Asn	gtg Val 295	gac Asp	cag Gln	gac Asp	ctc Leu	gtc Val 300	ggc	tgg Trp	caa Gln	gcg Ala	2713 ·
Pro					tcc											2761
305	Pro	GIY	ALG	Arg	310	Бец		110	СуБ	315	Cyb	GLY	per	Ser	320	

			agg Arg 340													2857
			tcg Ser													2905
ggc	atc Ile 370	ttt Phe	cgg Arg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cga Arg	gly aaa	gtt Val 380	gcg Ala	aag Lys	gcg Ala	gtg Val	2953
			ccc Pro													3001
			aac Asn													3049
			cac His 420													3097
			gca Ala													3145
			acc Thr													3193
			aac Asn													3241
			tac Tyr													3289
tct Ser	Gly 999	ggc	gcc Ala 500	Tyr	gac Asp	Ile	Ile	Ile	Cys	Asp	Glu	Cys	His	Ser	act Thr	3337
gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc Gly	atc Ile 520	ggc	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
acg	gct Ala 530	gga Gly	gcg Ala	cga Arg	ctc Leu	gtc Val 535	gtg Val	ctc Leu	gcc Ala	acc Thr	gct Ala 540	acg Thr	cct Pro	ccg Pro	gga Gly	3433
			gtg Val													3481

					ttt Phe											3529
					ctc Leu											3577
					ctg Leu											3625
					gta Val											3673
gtc Val 625	gta Val	gca Ala	acg Thr	gac Asp	gct Ala 630	cta Leu	atg Met	acg Thr	ggc Gly	ttt Phe 635	acc Thr	ggc	gat Asp	ttc Phe	gac Asp 640	3721
					aat Asn											3769
ctg Leu	gac Asp	ccg Pro	acc Thr 660	ttc Phe	acc Thr	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	gcg Ala	3817
gtg Val	tca Ser	cgc Arg 675	tcg Ser	cag Gln	cgg Arg	cga Arg	eso ejà aac	agg Arg	act Thr	ggt Gly	agg Arg	ggc Gly 685	agg Arg	atg Met	ggc	3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	Gly	atg Met	ttc Phe	gat Asp	3913
tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cys	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	Val	Arg	ttg Leu 730	Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009 ,
cca Pro	GJÀ aaa	ttg Leu	ccc Pro 740	gtc Val	tgc Cys	cag Gln	gac Asp	cat His 745	ctg Leu	gag Glu	ttc Phe	tgg Trp	gag Glu 750	ggc	gtc Val	4057
ttt Phe	aca Thr	ggc Gly 755	ctc Leu	acc Thr	cac His	ata Ile	gac Asp 760	gcc Ala	cat His	ttc Phe	ttg Leu	tcc Ser 765	cag Gln	act Thr	aag Lys	4105
					ttc Phe											4153
tgc Cys 785	gcc Ala	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	aag Lys 800	4201

tgt Cys	ctc Leu	ata Ile	cgg Arg	cta Leu 805	aag Lys	cct Pro	acg Thr	ctg Leu	cac His 810	gly aaa	cca Pro	acg Thr	ccc Pro	ctg Leu 815	ctg Leu	4249
					gtt Val											4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
					gta Val											4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc Gly	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
gga Gly	agg Arg	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	ttc Phe	4489
gat Asp	gag Glu	atg Met	gaa Glu 900	gag Glu	tgt Cys	gcc Ala	tca Ser	cac His 905	ctc Leu	cct Pro	tac Tyr	atc Ile	gaa Glu 910	cag Gln	gga Gly	4537
atg Met	cag Gln	ctc Leu 915	gcc Ala	gaa Glu	caa Gln	ttc Phe	aaa Lys 920	cag Gln	aag Lys	gca Ala	atc Ile	ggg Gly 925	ttg Leu	ctg Leu	caa Gln	4585
aca Thr	gcc Ala 930	acc Thr	aag Lys	caa Gln	gcg Ala	gag Glu 935	gct Ala	gct Ala	gct Ala	ccc Pro	gtg Val 940	gtg Val	gaa Glu	tcc Ser	aag Lys	4633
tgg Trp 945	cgg Arg	acc Thr	ctc Leu	gaa Glu	gcc Ala 950	ttc Phe	tgg Trp	gcg Ala	aag Lys	cat His 955	atg Met	tgg Trp	aat Asn	ttc Phe	atc Ile 960	4681
agc Ser	gjy aaa	ata Ile	caa Gln	tat Tyr 965	tta Leu	gca Ala	ggc	ttg Leu	tcc Ser 970	act Thr	ctg Leu	cct Pro	Gly	aac Asn 975	ccc Pro	4729
gcg Ala	ata Ile	gca Ala	tca Ser 980	ctg Leu	atg Met	gca Ala	ttc Phe	aca Thr 985	gcc Ala	tct Ser	atc Ile	acc Thr	agc Ser 990	ccg Pro	ctc Leu	47,77
acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	gjà aaa	gga Gly 1009	\mathtt{Trp}	gtg Val	gcc Ala	4825
gcc Ala	caa Gln 1010	Leu	gct Ala	cct Pro	ccc Pro	agc Ser 1015	Ala	gct Ala	tcc Ser	gct Ala	ttc Phe 1020	Val	ggc Gly	gcc Ala	gly ggc	4873

atc gct gga gcg g Ile Ala Gly Ala i 1025			Gly Lys Val	
gat att ttg gca g Asp Ile Leu Ala (
ttt aag gtc atg a Phe Lys Val Met s 1060	Ser Gly Glu Me			Val Asn
cta ctc cct gct a Leu Leu Pro Ala : 1075	Ile Leu Ser Pr			
tgc gca gcg ata (Cys Ala Ala Ile 1 1090	ctg cgt cgg ca Leu Arg Arg Hi 1095	c gtg ggc cca s Val Gly Pro	ggg gag ggg Gly Glu Gly 1100	gct gtg 5113 Ala Val
cag tgg atg aac o Gln Trp Met Asn 1 1105	cgg ctg ata go Arg Leu Ile Al 1110	g ttc gct tcg a Phe Ala Ser 1115	Arg Gly Asn	cac gtc 5161 His Val 1120
tee eee aeg cae t Ser Pro Thr His :				
cag atc ctc tct a Gln Ile Leu Ser 8 1140	agt ctt acc at Ser Leu Thr Il	c act cag ctg e Thr Gln Leu 1145	ctg aag agg Leu Lys Arg 1150	Leu His
cag tgg atc aac g Gln Trp Ile Asn (1155	Glu Asp Cys Se			
aga gat gtt tgg g Arg Asp Val Trp 2 1170				
tgg ctc cag tcc a Trp Leu Gln Ser I 1185			Gly Val Pro	
tca tgt caa cgt g Ser Cys Gln Arg G	ggg tac aag gg Gly Tyr Lys Gl 1205	a gtc tgg cgg y Val Trp Arg 1210	ggc gac ggc Gly Asp Gly	atc atg 5449 Ile Met 1215
caa acc acc tgc of Gln Thr Thr Cys I				Lys Asn
tgt tcc atg agg a Cys Ser Met Arg I 1235	Ile Val Gly Pr			

1250	att aac gcg Ile Asn Ala 125	Tyr Thr Thr			5593
ccg gcg cca aat Pro Ala Pro Asn 1265		Ala Leu Trp			5641
tac gtg gag gtt Tyr Val Glu Val			His Tyr Val		5689
acc act gac aac Thr Thr Asp Asn 130	Val Lys Cys		Val Pro Ala	-	5737
ttc aca gaa gtg Phe Thr Glu Val 1315				Pro Ala Cys	5785
aaa ccc ctc cta Lys Pro Leu Leu 1330		Val Thr Phe			5833
tac ccg gtt ggg Tyr Pro Val Gly 1345		Pro Cys Glu			5881
gtg ctc act tcc			cac att acg		5929
Val Leu Thr Ser	1365	Asp Pro Ser 1370	His Ile Thr	Ala Glu Thr . 1375	
gct aag-cgt agg Ala Lys Arg Arg 138	ctg gcc agg Leu Ala Arg	1370 gga tct ccc	ccc tcc ttg Pro Ser Leu	1375 gcc agc tca	5977
gct aag- cgt agg Ala Lys Arg Arg	ctg gcc agg Leu Ala Arg o ctg tct gcg	gga tct ccc Gly Ser Pro 1385 cct tcc ttg	ccc tcc ttg Pro Ser Leu	1375 gcc agc tca Ala Ser Ser 1390 tgc act acc Cys Thr Thr	5977 6025
gct aag-cgt agg Ala Lys Arg Arg 138 tca gct agc cag ser Ala Ser Gln	ctg gcc agg Leu Ala Arg ctg tct gcg Leu Ser Ala ccg gac gct	gga tct ccc Gly Ser Pro 1385 cct tcc ttg Pro Ser Leu 1400 gac ctc atc Asp Leu Ile	ccc tcc ttg Pro Ser Leu aag gca aca Lys Ala Thr 1405 gag gcc aac	gcc agc tca Ala Ser Ser 1390 tgc act acc Cys Thr Thr	
gct aag-cgt agg Ala Lys Arg Arg 138 tca gct agc cag Ser Ala Ser Gln 1395 cgt cat gac tcc Arg His Asp Ser	ctg gcc agg Leu Ala Arg ctg tct gcg Leu Ser Ala ccg gac gct Pro Asp Ala 141! ggc ggg aac	gga tct ccc Gly Ser Pro 1385 cct tcc ttg Pro Ser Leu 1400 gac ctc atc Asp Leu Ile 5	CCC tCC ttg Pro Ser Leu aag gca aca Lys Ala Thr 1405 gag gcc aac Glu Ala Asn 1420 gtg gag tca	gcc agc tca Ala Ser Ser 1390 tgc act acc Cys Thr Thr ctc ctg tgg Leu Leu Trp	6025
gct aag-cgt agg Ala Lys Arg Arg 138 tca gct agc cag Ser Ala Ser Gln 1395 cgt cat gac tcc Arg His Asp Ser 1410 cgg cag gag atg Arg Gln Glu Met	ctg gcc agg Leu Ala Arg ctg tct gcg Leu Ser Ala ccg gac gct Pro Asp Ala 141! ggc ggg aac Gly Gly Asn 1430 gac tct ttc	gga tct ccc Gly Ser Pro 1385 cct tcc ttg Pro Ser Leu 1400 gac ctc atc Asp Leu Ile 5 atc acc cgc Ile Thr Arg	CCC tCC ttg Pro Ser Leu aag gca aca Lys Ala Thr 1405 gag gcc aac Glu Ala Asn 1420 gtg gag tca Val Glu Ser 1435 caa gcg gag Gln Ala Glu	gcc agc tca Ala Ser Ser 1390 tgc act acc Cys Thr Thr ctc ctg tgg Leu Leu Trp gag aat aag Glu Asn Lys 1440 gag gat gag	6025 6073

1475	Met Pro			ccg gat Pro Asp		Pro I		
tta gag tcc Leu Glu Ser 1490			Asp Tyr					
tgt cca ttg Cys Pro Leu 1505	ccg cct Pro Pro	gcc aag Ala Lys 1510	gcc cct Ala Pro	ccg ata Pro Ile 1515	Pro Pro	cca o Pro 1	cgg agg Arg Arg 152	
aag agg acg Lys Arg Thr		Leu Ser				Ala I		
gag ctc gcc Glu Leu Ala				Ser Glu				
agc ggc acg Ser Gly Thr 1555	Ala Thr					Asp (
gcg gga tcc Ala Gly Ser 1570			Tyr Ser					
gag ccg ggg Glu Pro Gly 1585					Trp Ser			•
gag gag gct Glu Glu Ala		Asp Val				Tyr !		
	Ser Glu 1609 ctg atc	Asp Val acg cca	Val Cys	Cys Ser 1610 gcg gag Ala Glu	Met Ser gaa acc	Tyr :	Thr Trp 1615 ctg ccc	: 6697
Glu Glu Ala	Ser Glu 1609 ctg atc Leu Ile 1620 ctg agc Leu Ser	Asp Val acg cca Thr Pro aac tct Asn Ser	Val Cys tgc gct Cys Ala 1623	Cys Ser 1610 gcg gag Ala Glu 5 cgt cac Arg His	Met Ser gaa acc Glu Thr cac aac	aag of Lys 1 1630 ttg 9 Leu 1	Thr Trp 1615 ctg ccc Leu Pro	6697
Glu Glu Ala aca ggc gcc Thr Gly Ala atc aat gca Ile Asn Ala	ctg atc Leu Ile 1620 ctg agc Leu Ser	Asp Val acg cca Thr Pro aac tct Asn Ser agc gca	tgc gct Cys Ala 1629 ttg ctc Leu Leu 1640 agc ctg Ser Leu	Cys Ser 1610 gcg gag Ala Glu 5 cgt cac Arg His	Met Ser gaa acc Glu Thr cac aac His Asn 1649	aag (Lys 1630) ttg (Leu 1630)	Thr Trp 1615 ctg ccc Leu Pro gtc tat Val Tyr acc ttt	6745
Glu Glu Ala aca ggc gcc Thr Gly Ala atc aat gca Ile Asn Ala 1639 gct aca aca Ala Thr Thr	ctg atc Leu Ile 1620 ctg agc Leu Ser tct cgc Ser Arg	Asp Val acg cca Thr Pro aac tct Asn Ser agc gca Ser Ala 165	tgc gct Cys Ala 1629 ttg ctc Leu Leu 1640 agc ctg Ser Leu 5	Cys Ser 1610 gcg gag Ala Glu 5 cgt cac Arg His cgg cag Arg Gln tac cgg	gaa acc Glu Thr cac aac His Asn 1644 aag aag Lys Lys 1660 gac gtg Asp Val	aag (Lys 1630) ttg (Leu 1630) gtc (Val 1630)	Thr Trp 1615 ctg ccc Leu Pro gtc tat Val Tyr acc ttt Thr Phe	6745 6793

								91	93							
gaa Glu	gcc Ala	tgt Cys	aag Lys 170	Leu	acg Thr	ccc Pro	cca Pro	cat His 170	Ser	gcc Ala	aga Arg	tct Ser	aaa Lys 1710	Phe	ggc	6937
			Lys					Leu					gtt Val 5			6985
		Ser					Leu					Glu	aca Thr			7033
	Thr					Lys					Cys		caa Gln			7081
					Pro					Val			gat Asp		Gly	7129
gtt Val	cgt Arg	gtg Val	tgc Cys 1780	Glu	aaa Lys	atg Met	gcc Ala	ctt Leu 1789	Tyr	gat Asp	gtg Val	gtc Val	tcc Ser 1790	Thr	ctc Leu	7177
Pro	Gln	Ala 179	Val 5	Met	Gly	Ser	Ser 1800	Tyr)	Gly	Phe	Gln	Tyr 180		Pro	Gly	7225
Gln	cgg Arg 1810	Val	gag Glu	ttc Phe	ctg Leu	gtg Val 1815	Asn	gcc Ala	tgg Trp	aaa Lys	gcg Ala 1820	Lys	aaa Lys	tgc Cys	cct Pro	7273
atg Met 1825	Gly	ttc Phe	gca Ala	tat Tyr	gac Asp 1830	Thr	cgc Arg	tgt Cys	ttt Phe	gac Asp 1835	Ser	acg Thr	gtc Val	act Thr	gag Glu 1840	7321
aat Asn	gac Asp	atc Ile	cgt Arg	gtt Val 1849	Glu	gag Glu	tca Ser	atc Ile	tac Tyr 1850	${\tt Gln}$	tgt Cys	tgt Cys	gac Asp	ttg Leu 185	Ala	7369
				Gln					Leu				ctt Leu 1870	Tyr		7417
Gly ggg	ggc Gly	ccc Pro 187!	Leu	act Thr	aat Asn	Ser	aaa Lys 1880	Gly	cag Gln	aac Asn	tgc Cys	ggc Gly 1889	tat Tyr 5	cgc Arg	cgg Arg	7465
tgc Cys	cgc Arg 1890	Ala	agc Ser	ggt Gly	gta Val	ctg Leu 1895	Thr	acc Thr	agc Ser	tgc Cys	ggt Gly 1900	Asn	acc Thr	ctc Leu	aca Thr	7513
	Tyr					Ala					Ala		ctc Leu			7561

					Cys					Val.	gtt Val				Ser	7609
				Glu					Leu		gcc Ala			Glu		7657
atg Met	act Thr	aga Arg 1955	Tyr	tct Ser	gcc Ala	ccc Pro	cct Pro 1960	Gly	gac Asp	ccg Pro	ccc Pro	aaa Lys 1969	Pro	gaa Glu	tac Tyr	7705
gac Asp	ttg Leu 1970	Glu	ttg Leu	ata Ile	aca Thr	tca Ser 1975	Сув	tcc Ser	tcc Ser	aat Asn	gtg Val 1980	Ser	gtc Val	gcg Ala	cac His	7753
	Ala					Val					cgt Arg					7801
ccc Pro	ctt Leu	gcg Ala	cgg Arg	gct Ala 200!	Ala	tgg Trp	gag Glu	aca Thr	gct Ala 2010	Arg	cac His	act Thr	cca Pro	gtc Val 2015	Asn	7849
				Asn					Ala		acc Thr			Ala		7897
			Met					Ser			cta Leu		Gln			7945
		Lys					Gln				gcc Ala 2060	Cys				7993
Leu gag	Glu 2050 cca Pro	Lys) ctt	Ala gac	Leu cta	Asp cct	Cys 2055 cag Gln	Gln S atc	Ile att	Tyr	Gly cga	Ala 2060 ctc Leu	Cys) cac	Tyr ggc	Ser	Ile agc	7993 8041
gag Glu 2069	Glu 2050 cca Pro ttt	tca	gac Asp	Leu cta Leu cat	cct Pro 2070	Cys 2055 cag Gln)	Gln atc Ile tct Ser	Ile att Ile cca Pro	Tyr caa Gln ggt	cga Arg 2075 gag Glu	Ala 2060 ctc Leu	Cys cac His	Tyr ggc Gly agg	Ser ctt Leu gtg	agc Ser 2080 gct Ala	
gag Glu 2069 gca Ala	Glu 2050 cca Pro ttt Phe	tca Ser	Ala gac Asp ctc Leu	cta Leu cat His 2089	cct Pro 2070 agt Ser	Cys 2055 cag Gln tac Tyr	atc Ile tct Ser	att Ile cca Pro	caa Gln ggt Gly 2090 ccc Pro	cga Arg 2075 gag Glu	Ala 2060 ctc Leu atc	Cys cac His aat Asn	ggc Gly agg Arg	ctt Leu gtg Val 2095 aga Arg	agc Ser 2080 gct Ala	8041
gag Glu 2069 gca Ala tca ser	Glu 2050 cca Pro ttt Phe tgc Cys	tca Ser ctc Leu	gac Asp ctc Leu agg Arg 2100	cta Leu cat His 2089 aaa Lys	cct Pro 2070 agt Ser ctt Leu	cag Gln tac Tyr ggg Gly	atc Ile tct Ser gta Val	att Ile cca Pro ccg Pro 2109	caa Gln ggt Gly 2090 ccc Pro	cga Arg 2075 gag Glu ttg Leu	Ala 2060 ctc Leu 3 atc Ile	Cys cac His aat Asn gtc Val	gge Gly agg Arg tgg Trp 2110	ctt Leu gtg Val 2095 aga Arg	agc Ser 2080 gct Ala cat His	8041

	Leu					Ala					Asp			agc Ser	tgg Trp 2160	8281
	_	-			Ser			_		Tyr			_	tct Ser 2175	Arg	8329
_	-		_	Trp		_		-	Leu					gta Val O		8377
	_	_	Tyr			ccc Pro		Arg		acgg	gggag	gct a	aaac	actċo	ca	8427
tttt	tttt ctt	ett t	tttt ctag	tcct gtca	t tt	tttt	tect	ctt g aaa	tttt	tcc	ttt	tctt	cc f	tttgg	ttttt gtggct cagaga	8547

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 4 July 2002 (04.07.2002)

PCT

(10) International Publication Number WO 02/052015 A3

- (51) International Patent Classification7: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/257,857

22 December 2000 (22.12.2000) U

- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- (88) Date of publication of the international search report: 20 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A3

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.

International Application No PCT/CA 01/01843

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/51 C12N15/40 C12N7/04 C12N15/85

C12Q1/68

C12Q1/70

C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C12Q $\,$

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

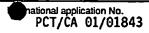
Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, MEDLINE, SEQUENCE SEARCH

C. DOCUM	ENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the	e relevant passages	Relevant to claim No.
X	LOHMANN V ET AL: "Replication subgenomic hepatitis C virus R hepatoma cell line." SCIENCE (WASHINGTON D C), vol. 285, no. 542, 2 July 1999 (1999-07-02), page XP002232924 ISSN: 0036-8075 the whole document	NAs in a	1-22
A	EP 1 043 399 A (BARTENSCHLAGER 11 October 2000 (2000-10-11) page 3 -page 24; tables 1,3	RALF DR)	1-22
X Furth	er documents are listed in the continuation of box C.	Patent family members are lister	d in annex.
'A" document consider of filing de filing de filing de current which is citation other m'P" document of the filing	nt which may throw doubts on priority claim(s) or s cited to establish the publication date of another or other special reason (as specified) nt referring to an oral disclosure, use, exhibition or	"T" later document published after the in or priority date and not in conflict wit cited to understand the principle or invention "X" document of particular relevance; the cannot be considered novel or cann involve an inventive step when the cannot be considered to involve an idecument is combined with one or ments, such combination being obvi in the art. "&" document member of the same pater	th the application but theory underlying the claimed invention of the considered to document is taken alone claimed invention inventive step when the nore other such docupous to a person skilled
	March 2003	Date of mailing of the international so	earch report
lame and m	ailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Schulz, R	

Interplication No PCT/CA 01/01843

	PCT/CA 01/01843
Chaudin of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
BLIGHT K J ET AL: "EFFICIENT INITIATION OF HCV RNA REPLICATION IN CELL CULTURE" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 290, 8 December 2000 (2000-12-08), pages 1972-1974, XP002951271 ISSN: 0036-8075 page 1972 -page 1973; table 1	1-22
	OF HCV RNA REPLICATION IN CELL CULTURE" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 290, 8 December 2000 (2000-12-08), pages 1972-1974, XP002951271 ISSN: 0036-8075



Box i Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
see additional sheet
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-22
Remark on Protest The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-22

A hepatitis C (HCV) replicon comprising: a 5'-non translated region (NTR) wherein guanin at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein further comprising one or more amino acid substitutions (adaptive mutations) in a non-structural protein and a 3'NTR; a eurkaryotic host cell transfected with said replicon; a RNA replication assay making use of said host cell and a method for testing compounds that inhibit HCV replication using said host cell.

2. Claims: 23-42 (in part)

A hepatitis C (HCV) replicon comprising: a 5'-non translated region (NTR), a HCV polyprotein region coding for a HCV polyprotein comprising a R(1135)K amino acid substitution (adaptive mutation) and a 3'NTR; a eurkaryotic host cell transfected with said replicon; a RNA replication assay making use of said host cell and a method for testing compounds that inhibit HCV replication using said host cell.

3. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(1148)G substitution.

4. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(1560)G substitution.

5. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a K(1691)R substitution.

6. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a L(1701)F substitution.

7. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a I $(1984)\,\mathrm{V}$ substitution.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

8. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a T(1993)A substitution.

9. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a G(2042)C or a G(2042)R substitution.

10. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(2404)P substitution.

11. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a L(2155)P substitution.

12. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a P(2166)L substitution.

13. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a M(2992)T substitution.

14. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a E(1202)G substitution.

ormation on patent family members

Intermenal Application No PCT/CA 01/01843

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
EP 1043399	A	11-10-2000	DE	19915178 A1	05-10-2000
			ΑT	236988 T	15-04-2003
			ΑU	2518000 A	19-10-2000
			CA	2303526 A1	03-10-2000
			DE	50001673 D1	15-05-2003
			EP	1043399 A2	11-10-2000
			JP	2001017187 A	23-01-2001